

## Zweisprachiger Schulunterricht für Migrantenkinder: Ergebnisse der Evaluationsforschung zu seinen Auswirkungen auf Zweitspracherwerb und Schulerfolg

Söhn, Janina

Veröffentlichungsversion / Published Version  
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
SSG Sozialwissenschaften, USB Köln

### Empfohlene Zitierung / Suggested Citation:

Söhn, J. (2005). *Zweisprachiger Schulunterricht für Migrantenkinder: Ergebnisse der Evaluationsforschung zu seinen Auswirkungen auf Zweitspracherwerb und Schulerfolg*. (AKI-Forschungsbilanz, 2). Berlin: Wissenschaftszentrum Berlin für Sozialforschung gGmbH FSP Zivilgesellschaft, Konflikte und Demokratie Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration -AKI-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-117016>

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Janina Söhn

**Zweisprachiger Schulunterricht  
für Migrantenkinder**

**Ergebnisse der Evaluationsforschung zu seinen  
Auswirkungen auf Zweitspracherwerb  
und Schulerfolg**

AKI-Forschungsbilanz 2

**Arbeitsstelle Interkulturelle Konflikte  
und gesellschaftliche Integration (AKI)  
Wissenschaftszentrum Berlin  
für Sozialforschung (WZB)**

**Juni 2005**



# **Zweisprachiger Schulunterricht für Migrantenkinder: Die AKI-Forschungsbilanz kurz gefasst**

## *Die AKI-Forschungsbilanzen*

Die AKI-Forschungsbilanz „Zweisprachiger Schulunterricht für Migrantenkinder. Ergebnisse der Evaluationsforschung zu seinen Auswirkungen auf Zweitspracherwerb und Schulerfolg“ erfasst und bewertet den gegenwärtigen Wissensstand zu diesem Thema. Absicht der AKI-Forschungsbilanzen ist es generell, den Kenntnisstand zu einer ausgewählten Thematik zu systematisieren und zu bewerten, hinreichend gesicherte Erkenntnisse von Hypothesen und kontroversen Positionen zu scheiden, Lücken und Desiderata aufzuweisen und damit Grundlagen für die Weiterentwicklung der Forschung und für politisches Handeln zu bieten.

## *Ausgangspunkt: unterdurchschnittlicher Bildungserfolg von Migrantenkindern*

Ausgangspunkt dieser Studie ist das unterdurchschnittliche Abschneiden von Kindern mit Migrationshintergrund an deutschen Schulen. Ein bedeutender Grund hierfür sind mangelnde Deutschkenntnisse. Viele dieser Kinder lernen in ihren Familien zunächst eine andere Sprache als Deutsch. Haben sie als Schulanfängerinnen und Schulanfänger noch kein muttersprachliches Niveau im Deutschen erreicht, besteht ohne gezielte Förderung das Risiko, dass sich dieses anfängliche Defizit über die gesamte Schulkarriere hinweg negativ auswirkt.

## *Zweisprachiger Unterricht: ein umstrittenes Modell*

Ein Modell einer solchen Förderung ist die bilinguale Erziehung. Im Kontext von Zuwanderung beinhaltet bilinguale Erziehung eine Alphabetisierung sowohl in der Muttersprache der Migrantinnen und Migranten als auch in der Landessprache; auch Fachunterricht wird in beiden Sprachen erteilt. Inwieweit Kinder mit Migrationshintergrund hiervon profitieren, ist umstritten. Zugunsten zweisprachiger Erziehung werden sprachwissenschaftliche und politische Argumente angeführt. Zum einen wird sie als förderlich für den Zweitspracherwerb und die kognitive Entwicklung erachtet. Zum anderen drückt sie die Anerkennung der Herkunftssprache als wichtige Ressource in einer kulturell und sprachlich vielfältigen Gesellschaft aus. Gegner bilingualer Unterrichtsmodelle hingegen befürchten primär, dass ein Unterricht in der Muttersprache nur auf Kosten des Zweitspracherwerbs möglich ist und letztlich gleichen Bildungschancen entgegensteht.

## *Fragestellung: Bilingualer Unterricht – Hindernis oder Basis für eine erfolgreiche Schullaufbahn?*

Dieser Bericht fragt danach, welchen Effekt zweisprachige Unterrichtsformen auf den Zweitspracherwerb und die Schulleistungen haben. Die Förderung der Zweisprachigkeit bzw. der Kompetenzen in der Erstsprache selbst steht hier nicht im Zentrum. Die Wirkung eines bilingualen Unterrichts wird im Vergleich zur Unterrichtung von Schülerinnen und Schülern derselben Herkunftssprache in einsprachigen Klassen (Regelunterricht oder spezielle Unterrichtung der Landessprache als Zweitsprache) beurteilt. Im Fokus stehen bilinguale Modelle an Grundschulen.

### *Empirische Basis: Auswertung vorhandener Evaluationsstudien*

Basis dieser AKI-Forschungsbilanz sind zwei neuere Metaanalysen sowie einzelne Studien, die bilingualen Unterricht für Migrantenkinder evaluierten. Die meisten dieser Forschungsarbeiten untersuchten spanisch-englische Unterrichtsmodelle in den USA. Sowohl die Metaanalysen selbst als auch ausgewählte Evaluationsstudien wurden vor allem mit einem methodenkritischen Fokus in Bezug auf das Untersuchungsdesign und die Datenanalyse ausgewertet. Unter diesen Gesichtspunkten erwiesen sich nur wenige Studien als aussagekräftig.

### *Hauptbefund: Wirkung auf Kompetenzerwerb in der Zweitsprache neutral*

Diese wenigen verlässlichen Studien lassen die Schlussfolgerung zu, dass zweisprachiger Unterricht im Vergleich zu monolingualen Alternativen in der Regel keine negative Wirkung auf die gemessenen Leistungen in der Zweitsprache hat. Ein konsistenter positiver Effekt kann nicht nachgewiesen werden. Speziell für die sprachwissenschaftliche Hypothese, dass Kinder zuerst in ihrer Erstsprache und erst darauf aufbauend in der Zweitsprache alphabetisiert werden sollten, finden sich bisher keine verlässlichen Belege.

### *Wenige Erkenntnisse über bilingualen Unterricht in Deutschland*

An deutschen Grundschulen gibt es nur wenige bilinguale Unterrichtsmodelle mit einer typischen Migrantensprache als Partnersprache neben dem Deutschen. Die seltenen wissenschaftlichen Begleitungen hierzu können aufgrund ihres Evaluationsdesigns bisher nicht die Frage beantworten, ob eine bilinguale Erziehung dem rein deutschsprachigen Unterricht überlegen ist.

### *Schlussfolgerung 1: Unterricht in der Erstsprache kein Hindernis für Zweitspracherwerb*

Auf Basis der US-amerikanischen Studien hat sich gezeigt, dass bilingualer Unterricht und hierbei insbesondere eine gleichzeitige Alphabetisierung in Erst- und Zweitsprache nicht auf Kosten des Zweitspracherwerbs gehen. Wird ein Teil der Unterrichtsstunden in einer anderen Sprache als der des Regelunterrichts erteilt, führt das offenbar nicht automatisch zu schlechteren Leistungen in der Zweitsprache und den Sachfächern. Wenn Kinder an bilingualen Schulprogrammen teilnehmen, ist also prinzipiell nicht zu befürchten, dass die so unterrichteten Schülerinnen und Schüler schlechter als solche in Regelklassen abschneiden werden.

### *Schlussfolgerung 2: kein Allheilmittel für Bildungsbenachteiligung*

Die uneinheitlichen Forschungsergebnisse zeigen, dass bilingualer Unterricht kein Allheilmittel zum Ausgleich von familiär bedingten schlechteren Lernausgangslagen etwa aufgrund eines bildungsfernen Elternhauses ist. Da aber eine Mehrheit der Migrantenkinder in Deutschland aus solchen Familien kommt, könnte eine Unterrichtung in ihrer Herkunftssprache nicht die Hauptmethode sein, um eine substanzielle Verbesserung ihrer Bildungskarrieren zu erreichen.

### *Forschungsbedarf zur Wirkung ein- und zweisprachiger Förderinstrumente*

Wie sich die Potenziale von Schülerinnen und Schüler mit nicht-deutscher Herkunftssprache am besten erschließen lassen, sollte eine zukünftige systematische Evaluation innovativer, miteinander konkurrierender deutsch- und zweisprachiger Förderprogramme klären helfen. Ein wichtiges methodisches Kriterium solcher Wirkungsanalysen sollte der Vergleich bilingual unterrichteter Kinder mit nur auf Deutsch unterrichteten Schülerinnen und Schülern der gleichen Herkunftssprachen sein. Ebenso wichtig ist die Erhebung und systematische Berücksichtigung weiterer individueller Schülermerkmale und von Charakteristika des Schulkontextes, die den Zweitspracherwerb und die Schulleistungen typischerweise beeinflussen. Wünschenswert ist eine stärker interdisziplinär kooperierende Forschung, die die methodischen und theoretischen Stärken von Pädagogik, Linguistik, Soziologie und Psychologie bündeln sollte.



## **Inhalt**

1. Einleitung	1
2. Spracherwerb, Zweisprachigkeit und Schule im Kontext von Migration – ein Problemaufriss	6
3. Bilinguale Unterrichtung: Pro und Contra	9
4. Bilinguale Schulprogramme	14
5. Ergebnisse und methodisches Vorgehen von Evaluationsstudien bilingualer Schulprogramme	19
5.1. Methodische Qualitätsmerkmale und Probleme	19
5.2. Ausgewählte US-amerikanische Evaluationsstudien: Kurzdarstellung und methodische Kritik	24
Ramirez u.a. (1991a, 1991b, 1991c)	25
Huzar (1973)	33
Alvarez (1975)	35
Thomas/Collier (1997 und 2002)	38
5.3 Studien aus Deutschland und anderen europäischen Ländern	46
Deutschland	46
Niederlande	49
Schweden	51
Exkurs: Studien über Kanada	51
5.4 Metaanalysen zur Effektivität zweisprachiger Schulprogramme	52
Greene (1998)	53
Slavin/Cheung (2005)	55
6. Überlegungen zu einer Evaluation von Modellversuchen zweisprachiger Unterrichtung in Deutschland	61
7. Resümee	64
Literatur	69





# 1. Einleitung<sup>1</sup>

Zweisprachige Erziehung ist eine von mehreren Optionen der schulischen Integration von Kindern, deren Muttersprache nicht die Verkehrssprache des Landes ist, in dem sie leben. Der Oberbegriff der bilingualen Erziehung „refers to the use of two (or more) languages of instruction at some point in a student’s school career. Each language is used as a medium of instruction to teach subject matter content rather than just the language itself.“ (Cummins 2003: 3)

In Deutschland ist diese Unterrichtsform für stärker verbreitete Zuwanderersprachen (z.B. Türkisch, Russisch, Polnisch, Spanisch, Italienisch, Portugiesisch, Griechisch) bisher nur in wenigen Modellversuchen umgesetzt worden. Für diese Migrantensprachen gibt es gut zwei Duzend bilinguale Grundschulen (Gogolin 2005: 137). Gängiger ist der freiwillige muttersprachliche (Ergänzungs-)Unterricht. Da hier aber nur wenige Stunden pro Woche eine Migrantensprache gelehrt wird und dies abgekoppelt vom sonstigen Regelunterricht meist unter pädagogisch unzureichenden Rahmenbedingungen geschieht, weicht diese sehr ‚abgespeckte‘ Form zweisprachiger Erziehung erheblich von deren Idealform ab. Üblich für Migrantenkinder in Deutschland ist das „Lernen[s] in Regelklassen mit ggf. zusätzlicher Förderung des Deutschen als Zweitsprache“ (Roth 2003: 379). In einer Einwanderungsgesellschaft wie den USA hat zweisprachige Erziehung dagegen eine längere Tradition, obwohl auch dort keineswegs alle Schülerinnen und Schüler mit nicht-englischer Muttersprache tatsächlich an zweisprachigen Programmen teilnehmen – 1993 waren dies etwa ein Zehntel aller als „limited English proficient“ eingestuften Kinder (Moran/Hakuta 2001: 446).

Politisch wie wissenschaftlich ist das Für-und-Wider zweisprachigen Unterrichts heiß umstritten (zur deutschen Debatte vgl. z.B. Hopf 2005; Brumlik 2000; Baur 2000; Reich 2000; Beck 1999; Kleff 1998; Zimmer 1998). Dabei geht es sowohl Kritikern als auch Befürwortern zweisprachiger Unterrichtsmodelle um die Verbesserung der schulischen Leistungen und darum, dass Kinder und Jugendliche unabhängig von ihrer Herkunftssprache die gleichen Chancen haben sollten, ihre Potenziale in schulische Erfolge umzumünzen.<sup>2</sup> Auch die Bedeutung der Verkehrssprache wird nicht grundsätzlich anders eingeschätzt: „Beide Auffassungen werden im Bewusstsein des Faktums vertreten, dass von der Beherrschung der Mehrheitssprache letztendlich der Bildungserfolg aller Kinder abhängt.“ (Fürstenau u.a. 2003: 20; vgl. auch Kultusministerkonferenz 2002: 15) Umstritten ist, ob den Zielen der guten Deutschkenntnisse und des bestmöglichen Schulerfolgs die Unterrichtung der und in der Muttersprache abträglich oder förderlich ist und unter welchen Bedingungen – wenn überhaupt – zweisprachiger Unterricht geeigneter als andere Unterrichtsmodelle ist, diese beiden Ziele zu erreichen. Gegenstand der Debatte ist daneben, welcher Wert der Förderung von Migrantensprachen zugemessen werden sollte. Unabhängig von einem ange-

---

<sup>1</sup> Ich bedanke mich bei Hartmut Esser, Karen Schönwälder, Christina Limbird, Friedhelm Neidhardt und Amélie Mummendey für ihre Kommentare und Anregungen beim Verfassen dieser AKI-Forschungsbilanz.

<sup>2</sup> In den 1960ern und 1970ern dagegen diente der Unterricht von „Gastarbeiterkindern“ in deren Muttersprache der Vorbereitung auf die erwartete Remigration in die Heimatländer und deren Schulsysteme. Diese Zielsetzung wird seit Anfang der 1980er nicht mehr prominent vertreten (vgl. Reich 2000: 114).

nommenen indirekten Nutzen kann dieser Kompetenzerwerb in der Muttersprache bzw. eine kompetente Zweisprachigkeit als ein eigenständiges, aus verschiedenen Gründen erstrebenswertes Bildungsgut angesehen werden: z.B. als wichtiger Bestandteil der eigenen kulturellen Identität und Bedingung der familiären Kommunikation, aber auch als nützliche ökonomische Ressource in einem zusammenwachsenden Europa und in internationalen Wirtschaftsbeziehungen. Umgekehrt gibt es Positionen, die eine Förderung der Erstsprache von Migranten als zu kostenträchtig oder als Beitrag zur Verfestigung unerwünschter ethnischer Zusammenhänge angesehen werden.

Zwei viel beachtete Gutachten nehmen in der deutschen Debatte relativ deutliche Positionen ein. Reich und Roth resümieren unter dem Vorbehalt verbleibender Forschungsdesiderata: „Unterricht im Medium der Erstsprache bewirkt nicht nur die weitere Entwicklung dieser Sprache, sondern auch höhere Leistungen im Medium der Zweitsprache (z. B. in Mathematik), wenn er über mehrere Jahre hinweg fortgeführt wird. Unterricht im Medium beider Sprachen ist ein starkes Instrument zur Verbesserung des Schulerfolgs zweisprachiger Schülerinnen und Schüler und kann, wenn weitere Qualitätskriterien erfüllt sind, zur Chancengleichheit mit einsprachigen Schülerinnen und Schülern führen.“ (2002: 42) Neben einer verbesserten Zweitsprachenförderung und anderen Maßnahmen betrachten es Gogolin, Neumann und Roth (2003) als „Erfolg versprechend, zweisprachig lebende Kinder schon im Elementarbereich in beiden Sprachen zu fördern. Insbesondere aber sollten sie in beiden Sprachen Zugang zur Schrift erhalten, und zwar in aufeinander abgestimmtem und ergänzendem Unterricht“ (Gogolin u.a. 2003: 61). In direkter Anlehnung an Gogolin gibt der Sachverständigenrat für Zuwanderung und Integration (2004: 263) eine vergleichbare Empfehlung und ähnlich befürwortet die Kultusministerkonferenz u.a. die „Entwicklung einer Didaktik der Mehrsprachigkeit für einen koordinierten Spracherwerb von Deutsch, Herkunftssprache und Fremdsprache(n) [und die] Förderung der vorhandenen Zwei- und Mehrsprachigkeit durch Weiterentwicklung des muttersprachlichen/herkunftssprachlichen Unterrichtsangebots u.a. im Hinblick auf eine stärkere inhaltliche und organisatorische Verzahnung mit dem Regelunterricht“ (Kultusministerkonferenz 2002: 16).

Allerdings ist in diesen Empfehlungen ebenso wie in dem im September 2004 angelaufenen Modellprogramm der Bund-Länder-Kommission (BLK) „Förderung von Kindern und Jugendlichen mit Migrationshintergrund FÖRMIG“ die Förderung der Muttersprache – im „Modul 6: Konzepte der Mehrsprachigkeit“ (BLK (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung) 2004: 10) – nur eine Fördermaßnahme unter vielen anderen.<sup>3</sup> Ob und in welcher Form Elemente zweisprachiger Erziehung in den einzelnen Bundesländern über Modellprojekte hinaus größere Verbreitung finden werden, ist derzeit noch ungewiss.

In der Kontroverse um bilingualen Unterricht ist auffällig, dass vorhandenes wissenschaftliches Wissen völlig gegensätzlich aufgefasst wird und konträre politische Folgerungen daraus abgeleitet werden. In einem neueren Bericht des schwedischen Skolverket, der nationa-

---

<sup>3</sup> In den bisher (Mai 2005) bewilligten Modellprojekten in Berlin, Brandenburg, Bremen, Hamburg, Nordrhein-Westfalen, Mecklenburg-Vorpommern, Rheinland-Pfalz, Saarland, Sachsen und Schleswig-Holstein nimmt die Förderung muttersprachlicher Kompetenzen jedoch keine herausragende Rolle ein, vgl. <http://www.blk-foermig.uni-hamburg.de/>.

len Bildungsagentur, zum muttersprachlichen Unterricht heißt es etwa: „All research points to the fact that mother tongue instruction increases the chances of reaching targets at school.“ (Skolverket 2002: 10) Folglich wird für eine Wiederbelebung des muttersprachlichen Unterrichts und der Förderung der Zweisprachigkeit von Migrantenkindern ab dem Kindergarten plädiert. Genau umgekehrte Schlüsse zog die niederländische Regierung aus der bisherigen Forschung. Der Bericht des parlamentarischen Untersuchungsausschusses zur Integrationspolitik resümiert: „...various scientific studies failed to prove the supporting function of OET (C) [Onderwijs in Eigen Taal (en Cultuur); Unterrichtung der familialen Herkunftssprache (und Herkunftskultur)].“ (Parliamentary Inquiry Committee on Integration Policy 2004: o.S.) Die niederländische Regierung hat im August 2004 diesen muttersprachlichen Unterricht abgeschafft. Weit über die Fachkreise von Bildungspolitikern und Pädagogen hinaus wurde das Thema der zweisprachigen Erziehung in den USA debattiert, als in mehreren Bundesstaaten in Referenden mehrheitlich für die Abschaffung der zweisprachigen Schulprogramme gestimmt wurde, so insbesondere 1998 in den Referenden ‚Proposition 227‘ in Kalifornien und im Jahr 2000 ‚Proposition 203‘ in Arizona.<sup>4</sup>

Diese internationalen Beispiele zeigen, wie unterschiedlich Nutzen und Ziele der muttersprachlichen Unterrichtung eingeschätzt werden können. Wenn es nun darum geht, solche schulischen bilingualen Programme und didaktische Maßnahmen neu zu entwickeln und zu implementieren bzw. sie zu reformieren und weiter zu fördern, ist es sinnvoll zu fragen, ob auch tatsächlich deren Ziele erreicht werden. Dieser Bericht will einen Beitrag dazu leisten, die bisherigen Erkenntnisse der Evaluationsforschung hierzu zu diskutieren und resümieren. Er untersucht (schwerpunktmäßig in Kapitel 5) die folgenden Hauptfragen:

**Was sagen bisherige Evaluationsstudien und Metaanalysen über die Effektivität zweisprachiger Schulprogramme bezogen auf die Kompetenzen in der Mehrheitsprache und die schulischen Leistungen aus?** Als Maßstab für Effektivität gilt der positive, neutrale oder negative Effekt des zweisprachigen Unterrichtsprogramms auf den Kompetenzerwerb und die relative Stärke des Effekts verglichen mit einer Kontrollgruppe von Schülern mit ähnlichem sprachlichem Hintergrund, die nur Unterricht in der Verkehrssprache hatten. Im Zentrum steht also die Frage, ob SchülerInnen bei einer Unterrichtung in der Mutter- und der Verkehrssprache oder nur in der Verkehrssprache in ihren Leistungen besser abschneiden. Evaluationsergebnisse zu darüber hinaus gehenden Differenzierungen zwischen unterschiedlichen bilingualen Programmen werden nur am Rande behandelt, da bei entsprechenden Studien meist der Vergleich zu den monolingualen Unterrichtsalternativen (z.B. in englischsprachigen Regel- oder Förderklassen) fehlt.

Die meisten einschlägigen Studien wurden in den USA durchgeführt, während es in Deutschland bisher keine Evaluationsstudien gibt, die die hier interessierende Frage systematisch beantworten könnten. In einem Unterkapitel (5.3) sollen die wenigen wissenschaftlichen Begleitforschungen zu zweisprachigen Modellversuchen in Deutschland und den europäischen Ländern Niederlande und Schweden kurz vorgestellt werden. Ausgangsbasis

---

<sup>4</sup> Als ein Beispiel für eine Pro-Zweisprachigkeits-Sicht der US-Debatte vgl. Lindholm-Leary (2001: 11-21), als Gegner zweisprachiger Programme s. Rossell (2003) und Imhoff (1990), vgl. auch den Themenband zu „Bilingualism and schooling in the United States“ im Journal for the Sociology of Language mit dem gleich lautenden Leitartikel von García (2002) sowie Heintze (2001).

für diesen Bericht waren ansonsten zwei Metaanalysen von Greene (1997; 1998)<sup>5</sup> bzw. Slavin und Cheung (2003; 2005)<sup>6</sup>. Neben diesen Metaanalysen werden exemplarisch drei der dort ausgewerteten Originalstudien (Ramirez u.a. 1991; Huzar 1973; Alvarez 1975) sowie die Forschungsarbeiten von Thomas und Collier (1997, 2002) näher vorgestellt und kritisch diskutiert (Kapitel 5.2). Insbesondere Greene (1998), Ramirez u.a. (1991) und Thomas/Collier (1997) dienen in den Gutachten von Reich und Roth (2002: 18-20) und z.T. von Gogolin, Neumann und Roth (2003: 45-49) als wichtige Belege für ihre Interpretation des Forschungsstands und die darauf basierenden Schlussfolgerungen.

**Wie wurden solche Evaluationsstudien und Metaanalysen methodisch durchgeführt und wie ist dies zu bewerten?** Eine Methodenkritik ist neben der Vorstellung zentraler inhaltlicher Ergebnisse der Originalstudien und Metaanalysen ein Schwerpunkt dieses Berichts. Referenzpunkte für diese Einschätzung sind die Merkmale von (Quasi-)Experimenten als Evaluierungsdesign (z.B. adäquate Auswahl einer Kontrollgruppe, Randomisierung, Berücksichtigung weiterer Einflussgrößen in multivariaten statistischen Verfahren), die auch in den Metaanalysen als Auswahlkriterien für methodisch akzeptable Studien dienen (Kapitel 5.1). Da Metaanalysen den Anspruch erheben, auf besonders systematische Weise die vorhandene Forschung zu resümieren, sollen die dort beschriebene Vorgehensweise und die erzielten Schlussfolgerungen kritisch betrachtet werden (Kapitel 5.4).

**Was wäre bei zukünftigen Evaluationsstudien im deutschen Kontext zu beachten und praktikabel?**

Basierend auf der Forschungserfahrung vor allem im US-Kontext, aber unter Berücksichtigung spezifisch deutscher Rahmenbedingungen, werden im Kapitel 6 in Ergänzung zu Kapitel 5.1 ein Design für eine Evaluationsstudie eines bilingualen Programms in Deutschland skizziert und methodische Überlegungen und Mindeststandards hierzu dargelegt.

Dem Hauptkapitel 5 zu den Evaluationsstudien vorangestellt sind ein kurzer Problemaufriß zum Zusammenhang von Spracherwerb, Zweisprachigkeit und Schulleistungen im Kontext von Einwanderungsgesellschaften (Kapitel 2) und anschließend eine Zusammenchau der in der Literatur vorgebrachten Argumente für und gegen zweisprachige Schulprogramme (Kapitel 3). Zudem wird in Kapitel 4 ein Überblick über unterschiedliche US-amerikanische Typen zweisprachiger Schulprogramme gegeben, auf die die Evaluationsstudien Bezug nehmen.

Als Zielgruppe sind in dieser Expertise primär zugewanderte Kinder oder Kinder zugewanderter Eltern von Interesse – und zwar solche, die in der Familie zunächst nicht oder nicht nur die Verkehrssprache des Aufnahmelandes (L2), sondern eine andere Sprache als Erstsprache (L1) sprechen lernen. Die Kinder, die also hier im Fokus stehen, sprechen bei der Einschulung entweder die Mehrheitssprache noch gar nicht oder sind in unterschiedlichem Maße zweisprachig. Für den Bildungsverlauf kann es vor allem problematisch sein,

---

<sup>5</sup> Der schon 1997 veröffentlichte Artikel von Greene unterscheidet sich inhaltlich nicht von der 1998 publizierten Version, die in diesem Beitrag verwendet wird.

<sup>6</sup> Die bisher zugängliche Publikation von 2003 wurde in der neueren Version von 2005 leicht modifiziert, wobei sich die Grundaussage nicht geändert hat.

wenn die L2-Kenntnisse bei der Einschulung noch nicht auf dem Niveau entwickelt sind, das für Muttersprachler in diesem Alter durchschnittlich zu erwarten ist. So hat etwa die PISA-Studie dargelegt, dass gerade eine andere Familiensprache einer der wichtigsten Faktoren zur Vorhersage von schwachen Lesekompetenzen war. Migrationsstatus, soziale Schicht, Einreisalter und die Umgangssprache in der Familie gehören zu den entscheidenden Faktoren, die die Leistungsdifferenz zwischen den untersuchten 15-jährigen SchülerInnen mit und ohne Migrationshintergrund erklären (Deutsches PISA-Konsortium 2001: 378f.). Es ist also davon auszugehen, dass für Kinder mit Migrationshintergrund die adäquate Beherrschung der Verkehrssprache eine notwendige, wenn auch nicht hinreichende Bedingung für gute schulische Leistungen ist. Um diese Sprachkompetenz zu fördern, können unterschiedliche pädagogisch-didaktische Konzepte und Unterrichtssprachen zum Einsatz kommen, die aber – und das ist der Schwerpunkt dieses Berichts – unterschiedlich effektiv sind.

Zweisprachige Programme können sehr unterschiedliche Schülergruppen ansprechen, bei denen sich die Rahmenbedingungen aber bedeutsam unterscheiden: Ausgeschlossen werden hier Programme für SchülerInnen der sprachlichen Majorität (oft Kinder aus mittleren oder gehobenen sozialen Schichten), die sich eine prestigeträchtige globale lingua franca wie Englisch oder Französisch durch zweisprachige Programme besonders intensiv aneignen sollen – in Deutschland die überwiegende Mehrheit der bilingualen Modelle (vgl. für einen Überblick über die Länderpolitiken Deutscher Bildungsserver 2004). Das freiwillige Erlernen einer weiteren Zweitsprache bleibt eine Bereicherung und keine Notwendigkeit wie bei EinwanderInnen. Denn alle gesellschaftliche Kommunikation sowie (größtenteils) die Bewertung der Schulleistungen geht weiterhin in der Muttersprache vonstatten (vgl. hierzu auch den Exkurs in Kapitel 5.3 zu bilingualen Schulmodellen in Kanada). Ebenfalls nicht berücksichtigt werden zweisprachige Schulen für Kinder aus angestammten nationalen bzw. hier auch linguistischen Minderheiten wie z.B. der sorbischen Minderheit in Deutschland. Diese Kinder erlernen heutzutage oft die Mehrheitssprache als Erstsprache, und das Erlernen der Minderheitensprache ist ein faktischer Zweitspracherwerb unter dem Vorzeichen der Pflege kulturell-sprachlicher Traditionen.

Insgesamt liegt der Fokus dieses Berichts auf bilingualen Erziehungsmodellen in der Primarstufe bzw. der Grundschule. Denn die Debatte darüber, ob eine Unterrichtung der Erstsprache für den weiteren Schulerfolg förderlich ist oder nicht, zielt vor allem auf die u.a. linguistisch begründete Bedeutsamkeit des Lesen- und Schreibenlernens in der Muttersprache und der Schulsprache, also auf die ersten Jahre des Schulbesuchs, ab.

## 2. Spracherwerb, Zweisprachigkeit und Schule im Kontext von Migration – ein Problemaufriss

Infolge von Migrationsprozessen kommt es dazu, dass die **Erstsprache (L1)** eines Kindes die Sprache der migrierten Eltern ist und die **Zweitsprache (L2)** üblicherweise gleichzusetzen ist mit der Landes- bzw. Verkehrs- und Unterrichtssprache in schulischen Regelklassen. Nur diese Fallkonstellation ist im Rahmen dieses Berichts von Interesse: Die Herausforderung der sprachlichen Integration stellt sich nicht, wenn dieselbe Sprache in Herkunfts- und Zielland gesprochen wird. Wenn nur ein Elternteil zugewandert ist und eine andere Muttersprache hat, also in bilingualen Partnerschaften, und die Kinder von Geburt an zwei Sprachen lernen, ist im Normalfall mit einer unproblematischen Aneignung beider Sprachen auf muttersprachlichem Niveau zu rechnen (Meisel 2004: 100; 2003: 33f).<sup>7</sup>

Migrantenkinder<sup>8</sup>, die im Aufnahmeland geboren oder als Kleinkinder zugewandert sind, können dabei im so genannten Primär- oder Erstspracherwerb unterschiedliche Ausprägungen einer „**muttersprachlichen (oder simultanen) Zweisprachigkeit**“ (Reich/Roth 2002: 11) entwickeln. Der **Primärspracherwerb** wird als die „Phase kindlicher Sprachaneignung, die mit dem ersten Lebenstag beginnt und allmählich endet, wenn eine Institution in den Aneignungsprozess steuernd eingreift“ (Gogolin u.a. 2003: 40) – also Kindergärten oder spätestens die Schule –, verstanden. Der **natürliche, ungesteuerte Spracherwerb** der Verkehrssprache in diesem Alter kann – falls nicht direkt durch die Eltern – etwa auch durch ältere Geschwister, Freunde oder Medienkonsum erfolgen und so zu unterschiedlichen Ausprägungen von Bilingualität führen. Im schulpflichtigen Alter zuwandernde Kinder (oder solche aus sprachlich völlig „isoliert“ lebenden Migrantenfamilien) werden dagegen eine so genannte **sukzessive Zweisprachigkeit** entwickeln. Beginnt ein Mensch in etwa zwischen dem fünften und zehnten Lebensjahr mit dem Erwerb von L2, spricht man von kindlichem, ab einem späteren Zeitpunkt von erwachsenem Zweitspracherwerb. Während der Schule und im Erwachsenenalter wird L2 durch natürlichen Spracherwerb ebenso wie – je nach Angebot und Wahrnehmung pädagogisch-didaktischer Unterstützung – durch gesteuerten Spracherwerb angeeignet. Unter Berücksichtigung erheblicher individueller Unterschiede sinkt die Wahrscheinlichkeit, in L2 ein muttersprachliches Niveau zu erreichen, je später die Aneignung dieser Zweitsprache anfängt (Meisel 2004: 103ff.; Butler/Hakuta 2004: 126ff.; Birdsong 1999).

Sowohl als Zwischenstufen als auch als (vorläufig) endgültige Entwicklungsstufen lassen sich unterschiedliche Kompetenzen in L1 und L2 feststellen. Idealtypisch lassen sich folgende Formen der Zweisprachigkeit ausmachen: „balanced“ oder **additiver Bilingualis-**

---

<sup>7</sup> Auch bei den deutschen PISA-Ergebnissen unterscheiden sich Jugendliche, bei denen beide oder ein Elternteil in Deutschland geboren sind, nicht signifikant (Deutsches PISA-Konsortium 2001: 378).

<sup>8</sup> Zur näheren Bezeichnung der hier interessierenden Personen: Hier im deutschen Kontext sind Kinder und Jugendliche mit Migrationshintergrund von Interesse, allerdings nur dann, wenn sie in ihrer Familie (auch) eine andere Sprache als Deutsch sprechen. Insofern kann man von zweisprachigen oder zweisprachig aufwachsenden Kindern (vgl. Reich/Roth 2002) sprechen, wobei hier begrifflich unklar bleibt, wann die Zweitsprache gelernt wird (je später, desto größer die pädagogische Herausforderung). In der US-amerikanischen Forschung spricht man sowohl von English Language Learner (ELL) oder Schülern mit Limited English Proficiency (LEP) als auch von Schülern einer Language Minority (LM), wobei hier nicht immer deutlich ist, auf welchem Niveau diese Kinder Englisch und die Zweitsprache tatsächlich sprechen.

**mus**, wenn hohe altersgemäße Kompetenzen in L1 und L2 erreicht und beibehalten werden; **dominanter Bilingualismus**, wenn nur eine der beiden Sprachen muttersprachlich beherrscht wird, und schließlich **Semilingualismus** bzw. **doppelte Halbsprachigkeit**, wenn in beiden Sprachen nur eine niedrige Kompetenz entwickelt ist (Siebert-Ott 1999: 30; Edwards 2004: 9f.; Rösch 2001: 28; Butler/Hakuta 2004: 115ff.; vgl. auch Skutnabb-Kangas/Toukomaa 1977).

Abb.1: Sprachliche Integration von Migranten

Kompetenz in Sprache des Herkunftslandes (L1)	Kompetenz in Sprache des Aufnahmelandes (L2)	
	ja	nein
ja	additiver Bilingualismus	monolingual in L1
	dominant bilingual in L1	
nein	dominant bilingual in L2	
	monolingual in L2	Semilingualismus

Je nach Alter bedeutet Kompetenz hierbei nur mündliche (Verstehen und Sprechen) oder zusätzlich schriftliche (Lesen und Schreiben) Fähigkeiten. So kann dominant bilingual bei einem Kindergartenkind bedeuten, L1 mündlich altersgemäß zu beherrschen und L2 erst gerade zu erwerben, bei einem Jugendlichen aber z.B. L2 altersgemäß in Schrift (biliteracy) und Wort zu beherrschen, während L1-Kenntnisse nur noch für eine informelle Konversation ausreicht.

Dieses Modell der Typen von Zweisprachigkeit kann man also auf die vier Komponenten von Sprachkompetenz anwenden (vgl. z.B. Edwards 2004: 8):

Abb. 2: Grundtypen von Sprachkompetenz

	schriftlich	mündlich
produktiv	Schreiben	Sprechen
rezeptiv	Lesen	Hörverstehen

Die Herausforderung für die lernenden Kinder und Jugendlichen ist es, sich während der Schulzeit durch das Schreiben- und Lesenlernen altersgemäß die „Schriftsprache“ (Maas/Mehlem 2003: 31) – bei Cummins (1979) „cognitive-academic language proficiency (CALP)“ – anzueignen, die „durch Merkmale konzeptioneller Schriftlichkeit geprägte Standardsprache/Fachsprache“ (Siebert-Ott 1999: 31). Diese ist von der „durch Merkmale



konzeptioneller Mündlichkeit geprägten Umgangssprache“ (Siebert-Ott 1999: 31) – bei Cummins (1979) „basic interpersonal communication skills (BICS)“ – zu unterscheiden.<sup>9</sup>

Schulen in Einwanderungsländern wie Deutschland oder den USA müssen bei Migrantenkindern der ersten und zweiten Zuwanderergeneration mit sukzessiver und simultaner Zweisprachigkeit umgehen. Bei Seiteneinsteigern ohne Kenntnisse der Unterrichtssprache liegt die pädagogische Herausforderung auf der Hand. Die Mehrheit der potenziell problematischen Fälle sind jedoch „zwei- oder mehrsprachig aufwachsende Kinder [...], von denen sich vor allem diejenigen in einer schwierigen Spracherwerbssituation befinden, die bei Schuleintritt die dominante Umgebungs- und Schulsprache nicht ausreichend beherrschen“ (Rösch 2001: 25). Gogolin und Kollegen betonen, dass im Vergleich zu einsprachig aufgewachsenen SchulanfängerInnen, bilingual aufgewachsene einen zwar nicht per se problematischeren, aber doch anderen Prozess der Sprachaneignung durchlaufen, da nur erstere auf eine „relative Homogenität der Quellen für Sprachaneignung“ (Gogolin u.a. 2003: 41) zurückgreifen können. Die daraus resultierenden Unterschiede im Sprachbesitz seien durch spezifische Fördermaßnahmen an Schulen zu berücksichtigen, damit sich im weiteren Verlauf aus diesen Differenzen nicht tatsächliche Nachteile entwickeln. Weniger die fließende mündliche Verständigung als die Aneignung der altersgemäßen Schriftsprache in L2 stellt für Kinder aus sprachlichen Minderheiten die nicht immer bewältigte Herausforderung dar (Cummins 1986: 149; Knapp 1999: 152).

Einerseits lässt sich wissenschaftlich beschreiben, welche Formen der Zweisprachigkeit sich bei Individuen und Gruppen tatsächlich herausbilden. Andererseits setzen pädagogische Maßnahmen und Interventionen normative Ziele, welche Sprachkompetenzen erreicht werden sollen. Unstrittig ist, dass ein Verharren in einem Semilingualismus weder für die Betroffenen noch die Schule oder die Gesellschaft insgesamt wünschenswert ist. Einsprachige Regelschulen sind primär für die Entwicklung und Leistungserbringung in der Unterrichts- und Verkehrssprache (also für Migrantenkinder in L2) verantwortlich und bleiben gegenüber der Entwicklung anderer Herkunftssprachen zumindest indifferent.<sup>10</sup> Eine Form des dominierenden Bilingualismus (etwa schriftliche und mündliche Beherrschung des Deutschen, nur mündliche Kenntnisse in der Migrantensprache) mag dann der normale outcome sein; die Entwicklung einer additiven Zweisprachigkeit bleibt der Privatinitiative der SchülerInnen bzw. ihrer Eltern überlassen. Für viele zweisprachige Schulmodelle dagegen ist die kompetente Zweisprachigkeit in Wort und Schrift ein Leitziel. Argumente für und gegen diese zweisprachige Unterrichtung werden im folgenden Kapitel vorgestellt.

---

<sup>9</sup> Diese eher grobe Unterteilung der Sprachkompetenzen von Cummins wurde von ihm später (Cummins 1986: 152f.) als zwei Kontinua „context-embedded/reduced communication“ einerseits und „cognitively undemanding/demanding communication“ andererseits rekonzeptionalisiert (vgl. aber auch neuere linguistische Ansätze der funktionalen Pragmatik z.B. bei Grieshaber 2001, 2004).

<sup>10</sup> Unbestritten gab und gibt es Beispiele, in denen die Muttersprache von Migranten und autochthonen Minderheiten aktiv unterdrückt und ihr Gebrauch im schulischen Kontext verboten wird.

### 3. Bilinguale Unterrichtung: Pro und Contra

Die im Folgenden vorgestellten Argumente für und gegen zweisprachige Erziehung bzw. die Unterrichtung der Herkunftssprache von MigrantInnen haben unterschiedliche Ausrichtungen: Es gibt gegensätzliche Hypothesen dazu, ob bilinguale Unterrichtung förderlich für den Erwerb der Zweitsprache und die schulische Leistungserbringung in der Zweitsprache ist. Diese Argumente werden jeweils als erste angeführt und stehen auch im Fokus dieses Berichts. Des Weiteren werden Argumente angeführt, die meist von einer unterschiedlichen Beurteilung des Wertes und des Nutzens von L1-Kompetenzen ausgehen und davon abgeleitet zugunsten oder gegen die Unterrichtung (in) der Erstsprache plädieren. Die Überlegungen rekurrieren teilweise auf linguistische, erziehungswissenschaftliche und soziologische Theorien, teilweise werden hier gesellschaftspolitische Zielstellungen formuliert.

#### Pro-Argumente

1. Das wichtigste sprachwissenschaftliche Argument zugunsten zweisprachiger Bildung betont die (angenommene) förderliche Wirkung der hohen Kompetenzen (insbesondere der Alphabetisierung) in L1 auf den Zweitspracherwerb. Die Interdependenz-Hypothese – auch als Transfer-Hypothese bekannt – von Jim Cummins postulierte hier folgenden Zusammenhang: „To the extent that instruction in Lx is effective in promoting proficiency in Lx, transfer of this proficiency to Ly will occur provided there is adequate exposure to Ly (either in school or environment) and adequate motivation to learn Ly” (Cummins 1981: 29). Diese Übertragung geschieht indirekt durch die Entwicklung einer verallgemeinerbaren kognitiven und linguistischen, mit Schriftlichkeit von Sprache zusammenhängenden Fähigkeit, die sich von einer Sprache auf die andere transferieren lässt. Nach Cummins (2003: 8f) geschieht diese Übertragung häufiger von der Sprache der (Migranten-) Minderheit auf die allgemeine Verkehrssprache (als bei Mitgliedern der Mehrheitsgesellschaft von der Verkehrssprache auf eine Zweitsprache), weil die Lernenden auch außerhalb der Schule häufiger mit Schriftlichkeit (literacy) in der Mehrheitssprache konfrontiert sind.

Cummins geht in seiner Schwellenniveau-Hypothese zugleich davon aus, dass für den Kompetenzerwerb von L2 entscheidend ist, auf welchem kognitiven Niveau L1 zu dem Zeitpunkt entwickelt ist, an dem Kinder intensiv mit der Zweitsprache konfrontiert werden, wie dies insbesondere in der formellen Beschulung (in L2) geschieht (Cummins 1982: 39). Die qualitative Abstufung von Kompetenzen in L1 und L2 bringt Cummins in der Schwellenniveau-Hypothese in Verbindung mit der kognitiven Entwicklung. „Das Erreichen eines niedrigen Schwellenniveaus zweisprachiger Kompetenz [d.h. einer dominanten Zweisprachigkeit] würde ausreichen, die negativen kognitiven Auswirkungen [des Semilingualismus] zu vermeiden, aber das Erreichen eines zweiten – höheren – Niveaus [d.h. additiver Bilingualismus] könnte die Voraussetzung für eine Beschleunigung kognitiver Entwicklung sein.“ (Cummins 1982: 38)

An diese Hypothesen anknüpfend postulieren etwa auch Thomas und Collier: „... if students do not reach a certain threshold in their first language, they may experience cognitive difficulties in the second language. [...] Furthermore, developing cognitively and linguisti-

cally in L1 at least throughout the elementary school years provides a knowledge base that transfers from L1 to L2.“ (1997: 41; vgl. auch Baur 2000: 131f.) Rösch (2001: 28) argumentiert, dass es nur eine spezielle, wenn auch gängige Auslegung der Interdependenzhypothese ist, dass die schulische Spracherziehung konsekutiv erfolgen sollte, während daraus auch eine gleichzeitige Alphabetisierung in L1 und L2 mit ähnlich positiven Transfereffekten abgeleitet werden kann.

2. Den inhaltlich-fachlichen Kompetenzerwerb stellt eine weitere Argumentationslinie von Thomas und Collier (2001) in den Mittelpunkt. Wird ein Kind eingeschult, das aus einem anderssprachigen Land neu zugewandert ist und keine oder kaum Kenntnisse in der Verkehrssprache des Aufnahmelandes besitzt, ist es ohne Fachunterricht in der Muttersprache zunächst mit dem Erlernen der Zweitsprache beschäftigt und kann kaum der Vermittlung inhaltlichen Wissens in der Regelklasse folgen: “Yet without grade-level academic work in primary language while acquiring English, these students get behind in their schooling and it is difficult for them to catch up to the constantly advancing native-English speakers who continue to make another year’s progress with each year of school.” (Thomas/Collier 2001: 282; vgl. auch Baker 2001: 235) Eine möglichst ununterbrochene kognitive Entwicklung eines Kindes soll demnach in der (besser beherrschten) Erstsprache erfolgen. Dieses Argument sollte auch für solche Zuwandererkinder der zweiten Generation gelten, die bis zur Einschulung kaum Kenntnisse in der Landessprache haben. Thomas und Collier nehmen nicht dazu Stellung, in welchem Maße dieses letzte Argument tatsächlich auch für Schulanfänger zutrifft, die durch ungesteuerten Erwerb der Zweitsprache die Sprache im Regelunterricht prinzipiell verstehen.

3. Aufgrund der innerfamiliären Kommunikationsmöglichkeit in L1 sprechen die Soziologen Mouw und Xie (1999) der additiven Zweisprachigkeit eine indirekt positive Wirkung auf den Schulerfolg zu: Solange die Migranteneltern nicht die Sprache des Aufnahmelandes sprechen, ist es wichtig, dass ihre Kinder immer noch die Herkunftssprache sprechen. Ohne die so gewährleistete Kommunikationsfähigkeit zwischen Kindern und Eltern „parents encounter difficulties in monitoring their children’s performance and instilling values and behaviors beneficial for academic achievement“ (Mouw/Xie 1999: 233). Nur mit diesem Argument der familiären Kommunikation seien zweisprachige Schulprogramme zu rechtfertigen. Diese Argumentation misst den mündlichen Kenntnissen in der Herkunftssprache die entscheidende Rolle zu. Inwiefern eine Alphabetisierung in L1 hier vonnöten ist, bleibt offen.

4. Ferner wird die psychologische und kommunikative Bedeutung der Erstsprache im familiären Zusammenhang unterstrichen: Muttersprachliche Unterrichtung unterstützt den Erhalt der Erstsprache, die seit der frühen Kindheit für die emotionalen Bindungen in der Familie prägend war (vgl. Tannenbaum 2003) und weiterhin für die Kommunikation innerhalb von Zuwandererfamilien bedeutend bleibt (Kleff 1998: 67f.; Bundesforum Familie 2004: 11; Portes/Hao 2002: 907). Durch eine Alphabetisierung in L1 wird zudem die schriftliche Kommunikation mit Familienmitgliedern in den Herkunftsländern möglich (Driessen 2005: 86). Dieses Pro-Argument ebenso wie die nun folgenden betont den Wert der Förderung der Erstsprache aus anderen Gründen als einen angenommenen indirekten Nutzen für den Zweitspracherwerb.

5. Die Unterrichtung in der Herkunftssprache drückt eine Anerkennung und Wertschätzung (durch das staatliche Schulsystem) der Migrantensprachen als gesellschaftliche Ressource und der sprachlich-kulturellen Identitäten von ZuwanderInnen aus und steuert so zu einem positiven Selbstbild von MigrantInnen bei (Baker 2001: 240; Portes/Hao 2002: 905; zur entsprechenden Kritik an monolingualen Schulsystemen vgl. Gogelin 2000: 88, 94). Als Teilaspekt der kulturellen Identität von Migranten ist durch L1-Kenntnisse in Wort und Schrift erst der Zugang und eine vertiefte Auseinandersetzung mit der (Hoch-)Kultur des Herkunftslandes möglich (Driessen 2005: 86). Diese Argumentationslinie ist meist eingebettet in eine breitere Perspektive, die kulturelle und sprachliche Pluralität in modernen Einwanderungsgesellschaften als positiven Wert betrachtet.

6. Der ökonomische Nutzen kompetenter Zweisprachigkeit ist schließlich ein weiteres wichtiges Argument zugunsten bilingualer Erziehung: In einem zusammenwachsenden Europa und einer globalisierten Wirtschaft ist die Förderung von Mehrsprachigkeit sowohl für die individuelle Erwerbstätigkeit z.B. in transnational agierenden Unternehmen und Organisationen als auch für eine Volkswirtschaft und ihre internationalen Handelsbeziehungen eine wichtige Investition (Kleff 1998: 68f.; Thomas/Collier 2002: 119). Um Mehrsprachigkeit als Ressource beruflich nutzen zu können, ist die Alphabetisierung in den Erstsprachen unerlässlich (Unabhängige Kommission „Zuwanderung“ 2001: 217). Die gilt auch für eine transnationale Lebensführung und berufliches Pendeln zwischen Herkunfts- und Aufnahmeland (vgl. Gogolin/Pries 2004: 16).

### **Contra-Argumente**

1. Das Hauptargument gegen bilinguale Erziehung bzw. muttersprachlichen Unterricht ist, dass der Unterricht (in) der Herkunftssprache nur auf Kosten des Erlernens der Landessprache und damit der Schulleistungen insgesamt möglich ist (vgl. Hopf 2005: 241f.; kritisch hierzu Reich 2000: 113). Rossell als eine der profiliertesten Kritikerinnen zweisprachiger Schulprogramme sieht in ihrer Zusammenschau bisheriger Forschungsevidenz folgenden Zusammenhang: „The highest level of proficiency in a second language is attained by being instructed in that language from the beginning of one’s education.” (Rossell 1990: 80) Für sie ist das Prinzip “time on task” (als bedeutende Komponente im Modell schulischen Lernens von J.B. Carroll (1963) eingeführt, vgl. hierzu Hopf 2005: 242f.) das entscheidende, wie lange also Kinder einer Lern- und Kommunikationssituation in L2 exponiert sind: “The amount of time spent learning a subject is the greatest predictor of achievement in that subject.” (Rossell 1990: 81; vgl. auch Imhoff 1990: 51; für Deutschland vgl. die Kritik am muttersprachlichen Unterricht von Brumlik 2000; sowie Zimmer 1998).

2. Cummins Interdependenz- und Schwellenhypothese wurde durch die bisherige empirische linguistische Forschung nur teilweise bestätigt. Verhoeven (1994: 408f.) zeigt in seiner empirischen Studie, dass Transfer von L1 auf L2 sich insbesondere bei pragmatischen Fähigkeiten, aber nicht in allen sprachlichen Teilbereichen nachweisen lässt und zudem Transfer in beide Richtungen möglich ist, d.h. dass schriftsprachliche Kompetenzen auch umgekehrt von L2 auf L1 übertragen werden können (vgl. auch Maas u.a. 2004: 139). Vor allem ist die eigenständige Wirkung (zwei-)sprachlicher Entwicklung auf Kognition umstritten (Edwards 2004: 15ff.; Portes/Rumbaut 2001: 117; Verhoeven 1987: 80ff.; Hopf 2005:

240; Bialystok 2004). Die kausale Notwendigkeit der Alphabetisierung in L1, wenn auch nicht die vorhandene Korrelation zwischen L1- und L2-Kompetenzen, wird so in Frage gestellt.

3. Zum Teil wird für Kinder, die sich L2 nicht mehr im Rahmen des Primärspracherwerbs als Kleinkinder aneignen, die bilinguale Unterrichtung als ungeeignet eingeschätzt: Abgeleitet von der „Interlanguagehypothese“ (Selinker 1992) zum Zweitspracherwerb wird ein Zweitsprachunterricht empfohlen, der sich am natürlichen, ungesteuerten Zweitspracherwerb orientiert und „weitgehend ohne Bezug zur Erstsprache“ (in Bezug auf Felix (1982) vgl. Rösch 2001: 28; vgl. auch Kupfer-Schreiner 1994: 41ff.) bleibt. In der deutschen Debatte hat etwa Felix die These aufgestellt, dass die Erkenntnisse kanadischer Immersion-Programme (s. Exkurs, S. 51) „die L2-einsprachige Erziehung als günstiger erscheinen lässt als den Gebrauch zweier Unterrichtssprachen“ (Felix 1993: 302 in Siebert-Ott 1999: 45).

4. Der potentielle Erfolg zweisprachiger Erziehung hängt von der linguistischen Distanz der Partnersprachen ab. So ist etwa bei einem nicht-romanischen Alphabet oder allgemein bei sprachlich nicht-verwandten Sprachen, z.B. dem Hoch-Arabischen und einer germanischen Sprache, fraglich, ob es bei einer gleichzeitigen Alphabetisierung nicht eher zu negativen Interferenzen kommt bzw. zumindest keine Transferleistung im oben genannten Sinne wahrscheinlich ist (Driessen 2005: 84; Rossell 2003: 47).

5. Wenn im L1-Unterricht faktisch die Amtssprache des Herkunftslandes unterrichtet wird, obwohl die Kinder einen von der Hochsprache stark divergierenden Dialekt (wie im Arabischen oder Mandarin) oder die Sprache einer Sprachminderheit (wie Kurdisch oder Berber) sprechen, ist kein positiver Übertragungseffekt im Sinne von Cummins Interdependenz-These zu erwarten (Rossell 2003: 46; Driessen 2005: 83f.). Anstelle der Muttersprache wird hier eine weitere Fremdsprache erlernt (wie es in den Niederlanden bei Kindern mit Berber als Erstsprache im Arabischunterricht in erheblichem Umfang stattfand).

6. Je nach Organisation von Regel- und muttersprachlichem Unterricht kann die Vermittlung von L1-Kenntnissen auch auf Kosten von Fachwissen geschehen. Zumindest für den niederländischen Fall, wo muttersprachlicher Unterricht während des Regelunterrichts bis zu 2,5 Stunden am Vormittag stattfand und so Sachunterricht auf Niederländisch verpasst werden konnte, konnte dies zu entsprechenden fachlichen Defiziten führen, die schwer aufzuholen waren (Driessen 2005: 84).

7. Aus einer deutlich normativen Perspektive sprechen sich Kritiker gegen solche Arten bilingualer Programme aus, die Kinder aus Minderheiten in eigenen (bilingual unterrichteten) Klassen und damit sozial segregieren (Imhoff 1990: 57). In einer pluralistischen Gesellschaft „government should foster the similarities that unite us rather than the differences that separate us“ (Imhoff 1990: 49; politische Diskussionen in Deutschland hierzu referiert Reich 2000: 118ff.).

8. Das Argument der ökonomischen und beruflichen Nützlichkeit hoher L1-Kompetenzen von Personen mit Migrationshintergrund gilt insofern nicht uneingeschränkt, als die „Verwertbarkeit“ und Nachfrage nach bestimmten Sprachkenntnissen auf dem Arbeitsmarkt und in Wirtschaftszweigen unterschiedlich ausgeprägt ist (zum unterschiedlichen Gebrauchswert von Sprachen in mehrsprachigen Kontexten vgl. allgemein de Swaan 2001; zur Rolle

von Migrantensprachen bei der beruflichen Integration vgl. Grin u.a. 2003). Dies ist allerdings nur ein Argument dafür, dass es etwa in politischen Entscheidungen abzuwägen gilt, wie vielen Kindern welcher Muttersprache eine intensive bilinguale Erziehung ermöglicht bzw. hierfür die Finanzierung bereitgestellt werden sollte.

Im Zentrum dieses Berichts stehen die Pro- und Contra-Argumente, die sich auf die Auswirkungen bilingualer Erziehung auf den Zweitspracherwerb und Schulleistungen in L2 beziehen. Dies ist auch der Fokus der meisten Evaluationsstudien zu solchen Unterrichtsmodellen. Daher soll nicht diskutiert werden, ob Mehrsprachigkeit per se (also unabhängig von Unterrichtsmodellen) einen Einfluss auf die Intelligenz eines Kindes hat und mit besseren oder schlechteren Schulleistungen assoziiert ist (vgl. hierzu zusammenfassend Moran/Hakuta 2001: 447-449). Ebenso wenig sind divergierende Thesen zur Bedeutsamkeit guter L1-Kompetenzen für das Selbstwertgefühl, die familiäre Kommunikation oder ihr beruflich-ökonomischer Nutzen weiter Gegenstand dieses Berichts, da auch der direkte Zusammenhang mit muttersprachlicher Unterrichtung schwieriger nachzuweisen wäre. Gefragt wird im Folgenden also nach der Auswirkung zweisprachiger Unterrichtung auf den L2-Erwerb und den schulischen Erfolg (unter Berücksichtigung anderer möglicher Einflüsse). Zunächst werden aber unterschiedliche Varianten bilingualer Unterrichtsmodelle vorgestellt.

## 4. Bilinguale Schulprogramme

Im Folgenden sollen unterschiedliche Typen zweisprachiger Schulprogramme vorgestellt werden, wie sie insbesondere in den USA und Kanada entwickelt und evaluiert wurden. Der folgende Überblick stützt sich auf die jeweils etwas variierenden zusammenfassenden Darstellungen bei Reich/Roth (2002: 17-18, 21), Siebert-Ott (1999: 20), Thomas/Collier (1997: 58) sowie Verhoeven (1987: 11). Die Angaben zu der anteiligen Unterrichtung in L1 und L2 sind für alle Modelle so zu verstehen, dass jeweils ganze Unterrichtsstunden bzw. Fächer in einer der beiden Sprachen unterrichtet werden, und nicht fortwährend, insbesondere nicht durch ein und dieselbe Lehrerin, zwischen den Sprachen hin und her gewechselt wird.

**Bilingual Immersion Education/Dual Language Education (DLE)** bzw. in Deutschland bilinguale Modellschulen/-klassen:

- Fachunterricht wird idealerweise während der gesamten Schulzeit in L1 und L2 erteilt.
- Im **90-10-Modell** wird in den ersten beiden Jahren der Beschulung<sup>11</sup> zu 90% Unterricht in der Muttersprache erteilt; in den letzten 5 bis 6 Schuljahren erreicht der Anteil in L2 60%. Im **50-50-Modell** hält sich die Unterrichtung in beiden Sprachen die Waage; L2 wird in den letzten 5 bis 6 Jahrgangsstufen auf 60% gesteigert.
- Dieses zweisprachige Modell kann sich an Schüler einer Sprachgruppe („**one way**“) oder zweier Sprachgruppen („**two way**“) wenden, wobei eine Schülergruppe jeweils die Muttersprache der anderen erlernt (z.B. die normale Verkehrssprache und eine Migrantensprache) und die informelle Kommunikation mit Muttersprachlern innerhalb einer Klasse den Spracherwerb noch zusätzlich unterstützt.
- Bei diesen „starken“ zweisprachigen Programmen kann schließlich zwischen konsekutiver oder gleichzeitiger Alphabetisierung unterschieden werden, d.h. ob Kinder zunächst in L1 alphabetisiert werden, während sie sich L2 mündlich aneignen, oder ob sie gleichzeitig und von Anfang an in L1 und L2 Lesen und Schreiben lernen (Slavin/Cheung 2003: 2; Lindholm-Leary 2001: 36).

Rösch unterscheidet bei einer gleichzeitigen Alphabetisierung in L1 und L2 noch „koordinierte“ Konzepte, bei denen „beide Sprachen weitgehend unabhängig von einander erworben und verwendet werden“, von einem Modell der „kombinierten“ Zweisprachigkeit, bei dem im Unterricht versucht wird, „eine Verbindung zwischen beiden Sprachen herzustellen, um Interferenzen zu vermeiden, Sprachvergleiche anzugeben“ (2001: 30) und damit vor allem eine drohende doppelte Halbsprachigkeit verhindert werden soll. Soweit Evaluationsstudien zu Schulprogrammen überhaupt genügend Informationen liefern, scheinen US-Modelle der gleichzeitigen Alphabetisierung in der Mehrheit „koordiniert“ vorzugehen.

- Im Gegensatz zu diesen Modellen, wie sie in und aus der nordamerikanischen Literatur kommend rezipiert werden, führt Verhoeven (1987: 11) für den niederländischen Fall bei den konsekutiven Modellen – er nennt sie sukzessiv – die L2-L1-

---

<sup>11</sup> In den USA beginnt nach einem 1-bis-2-jährigen Besuch einer nursery school (dt. eher Kindergarten) die vorschulische Unterrichtung mit einem ein- bis zweijährigen Kindergarten-Besuch (K) (dt. eher Vorschule); auf die Grundschule, meist bis zur 6. Klasse, folgen Junior (7.-9. Klasse) und Senior High School (10. bis 12. Klasse) (U.S. Department of Education 2004).

Option auf, die zuerst (im Regelunterricht) in der Zweitsprache, und zu einem späteren Zeitpunkt (etwa ab der zweiten Klasse im muttersprachlichen Ergänzungsunterricht) in der Erstsprache alphabetisiert. Dabei steht diese Reihenfolge allerdings im Widerspruch zu Cummins Schwellenniveauhypothese.

### **Developmental (DBE)/Maintenance/Late-Exit (Transitional) Bilingual Education (TBE):**

- Bei diesem Modell ist die Zahl der Schuljahre, in denen Unterricht in L1 und L2 erteilt wird, begrenzt: In den ersten (Vor-)Schuljahren dominiert L1 (ca. 80%), bis zur 6. Klasse (Ende der Grundschule) dann L2 (L1 ca. 20%). Danach erfolgt der Übergang in die monolingualen Regelklassen. Ansonsten gibt es prinzipiell auch hier die one- und two-way-Varianten.

In den bisher genannten Modellen sind neben der Beherrschung der Zweitsprache gute, altersgerechte Kenntnisse in L1, also eine additive Zweisprachigkeit, ein eigenständiges Ziel pädagogischen Wirkens. Deshalb werden sie auch als „enrichment models“ (Thomas/Collier 1997: 59) bezeichnet.

### **Early-Exit (Transitional) Bilingual Education/Transitorisches Modell**

- In diesem Modell ist dagegen die zeitweilige Unterrichtung und Alphabetisierung in L1 vor allem Mittel zum Zweck des besseren Erlernens der Verkehrssprache, bzw. der Leistungserbringung in dieser Sprache. Fachunterricht erfolgt zunächst (d.h. meist Vorschulklasse und erste Jahrgangsstufe) in L1 und L2. Innerhalb von zwei bis drei Jahren (und damit wesentlich rascher also bei den oben genannten Modellen) erfolgt ein allmählicher Übergang zum 100%-igen Unterricht in der Zweitsprache.

In Abbildung 3 sind die wichtigsten Eigenschaften dieser unterschiedlichen nordamerikanischen Modelle, die in der Literatur leider verwirrend vielfältig bezeichnet werden, inklusive monolingualer Alternativen der Förderung des Zweitspracherwerbs zusammengestellt. Während schon die Beschreibung der theoretischen Modelle zwischen verschiedenen Autoren leicht variieren, so werden die eigentlichen Implementierungen trotz gleicher Label oft sehr variantenreich ausfallen.

Diesem Schema lassen sich Modelle aus europäischen Ländern nur bedingt zuordnen. Muttersprachlicher Unterricht bzw. muttersprachlicher Ergänzungsunterricht in Deutschland, in den Niederlanden (Onderwijs in Allochthone Levende Talen – OALT) und Schweden (Hemspråkundervisning) können zwar in etwa der Developmental/Maintenance Bilingual Education zugerechnet werden (vgl. Siebert-Ott 1999: 21).<sup>12</sup> Die Vergleichbarkeit mit diesem US-amerikanischen Modell hängt aber von der Intensität (wöchentlicher Umfang) und Dauer (in Jahren) ab, in der dieser Unterricht tatsächlich angeboten und besucht wird. Zudem ist Sachunterricht in L1 kein üblicher Bestandteil dieses „europäischen“ muttersprachlichen Unterrichts.

---

<sup>12</sup> Vergleichbar mit transitorischen Modellen, in denen auch Sachunterricht in L1 erteilt wird, erscheinen die zweisprachigen Klassen für ausländische Kinder in Bayern, die insbesondere in den 1980ern eine verbreitete Unterrichtsform waren (vgl. Nitzschke 1982: 129); erst Mitte der 1990er sank der Anteil von ausländischen SchülerInnen, die diese Klassen anstelle des Regelunterrichts besuchten auf unter 10 % (Hunger/Thränhardt 2004: 193; kritisch hierzu Beck 1999: 8f.).



Abb. 3: Modelle der zweisprachigen Erziehung und monolinguale Alternativen

Unterrichtsmodell ggf. mit Varianten	Angestrebte Sprachkompetenz	Alphabetisierung	Umfang der Unterrich- tung in L1	Dauer des Modells in Jahrgangsstufen	Unterrichtung mit Schülern der Majoritätssprache
Bilingual Immersion Education/ Dual Language Education	L1 und L2	drei Varianten: ▪ konsekutiv: erst in L1, dann L2; ▪ konsekutiv: erst in L2, dann L1 ▪ koordiniert: gleich- zeitig in L1 + L2			
90-10-Modell			Ca. 90% in den ersten beiden Schuljahren, bis 12. Klasse 40%	alle Jahrgangsstufen	
50-50-Modell			bis 6. Klasse 50%, da- nach 40%	alle Jahrgangsstufen	
two-way					ja (ideal: 50:50)
one-way					nein
Developmental /Maintenance/ Late- Exit (Transitional) Bilingual Education	L1 und L2	konsekutiv: zunächst in L1	von ca. 80% in Vorschu- le bis 20% in 6. Klasse	Vorschuljahr bis 5. oder 6. Klasse	
two-way					ja (ideal: 50:50)
one-way					nein
Early-Exit Transitional Bilingual Edu- cation (Transitorisches Modell)	L2 (mit anfänglicher Alphabetisierung in L1 als Unterstützung)	konsekutiv: zunächst in L1	von 33-50% in Vorschu- le bis 0% in 4. Klasse	Vorschuljahr bis Ende 1. oder 2. Klasse, max. 3. Klasse	nein
Systematische Förderung des Zweit- spracherwerbs	L2	L2	0%, aber informell spon- tanes Übersetzen durch ggf. bilinguale Lehrer möglich	Primar- und Sekundar- stufe	
Structured Immersion/ Sheltered Instruction /~ Vorbereitungsklassen (Sprach- und Sachunterricht)					z.T.: variiert zw. 50 % und 0% pro Schultag
Pullout/taught as a subject/ ~ Deutsch-Förderkurse (nur Sprachunterricht)					ja, bis auf 30 Min. bis 2 Unterrichtsstunden separa- ter Unterricht pro Tag
Submersion/”sink or swim” ~ Regel- unterricht ohne spezielle Förderung	L2	L2	0%	alle Jahrgangsstufen	ja

In Deutschland unterscheiden sich die Regelungen und Praktiken des muttersprachlichen Ergänzungsunterrichts von Bundesland zu Bundesland. Muttersprachlicher Unterricht wird zusätzlich zum Regelunterricht oft nachmittags erteilt, umfasst zwei bis fünf Wochenstunden, ist meist freiwillig und nicht mit anderen Fächern integriert. Zum Teil liegt der muttersprachliche Unterricht ganz in staatlicher Verantwortung (Bayern, Hessen, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz), zum Teil wird er in Verantwortung der Konsulate durchgeführt. In mehreren Bundesländern, kann anstelle der zweiten (in selteneren Fällen der ersten) Fremdsprache Unterricht in der Muttersprache bis hin zur Abiturprüfung gewählt werden (Gogolin 2005: 136f.; Gogolin u.a. 2001a; und diese Studie zusammenfassend Baumann 2001; Broeder/Extra 1999: 82-87; Allemann-Ghionda 1999: 68ff.; Büchel/Bühler-Otten 2000).<sup>13</sup>

Die Unterrichtung von und in L1 (Alphabetisierung und Sachunterricht in der Muttersprache) ist eine – variantenreiche – pädagogische Option mit dem Ziel, die unterschiedlichen Lernvoraussetzungen von Kindern mit noch nicht ausreichenden Kenntnissen in der Verkehrssprache zu berücksichtigen. Eine andere Alternative ist die systematische Förderung des Zweitspracherwerbs ohne Bezug auf die Erstsprache (vgl. auch Rösch 2001: 29f), wobei im einsprachigen Unterricht in der Verkehrssprache didaktische Praktiken des Lehrens der Unterrichtssprache als Zweitsprache eingesetzt werden: **English as a Second Language (ESL)**, **English to Speakers of Other Languages (ESOL)** bzw. **Deutsch als Zweitsprache (DAZ)**. Dabei ist eine wesentliche Unterscheidung, ob L2 nur in Form von Sprachunterricht geschieht, oder auch in Sachunterricht ESL/DAZ-Methoden zum Einsatz kommen (s. auch Tabelle). In Structured-Immersion-Klassen etwa wird meist für einen begrenzten Zeitraum von (idealerweise) 1 bis 2 Jahren in separaten Klassen intensiv die Zweitsprache gelehrt, wobei z.B. musische Fächer und Sport zusammen mit MitschülerInnen der Regelklassen besucht werden können. Während in den USA solche Programme allgemein für English Language Learner (ELL) gedacht sind, entsprechen diesem Modelltyp in Deutschland am ehesten die Vorbereitungsklassen z.B. in Baden-Württemberg oder Hamburg bzw. Sprachlernklassen in Bayern, die sich hauptsächlich an neu zugewanderte Kinder, insbesondere Aussiedlerkinder, wenden, die als Seiteneinsteiger kaum Kenntnisse in der Landessprache haben (Zickgraf 2005). In etwa ließen sich unter diese Kategorien der Structured Immersion auch die Intensivdeutschkurse für Kinder im Jahr vor der Einschulung zuordnen, die zusammen mit der derzeitigen Einführung von Verfahren der Sprachstandsfeststellung in mehreren Bundesländern im Kommen sind (a.a.O.). „Pullout“-Kurse entsprechen z.B. den Deutsch-Förderkursen in Hessen, wöchentlich zwei zusätzliche Stunden Deutschunterricht für Kinder mit entsprechendem Förderbedarf.

Schließlich existiert noch die Variante, dass den Kindern mit anderer Herkunftssprache gar keine gesonderten Angebote gemacht, sondern sie von vornherein in einsprachige Regelklassen ohne spezielle Förderung eingeschult werden. Im Amerikanischen wird diese Alter-

---

<sup>13</sup> M.E. existieren keine bundesweiten Übersichten aus amtlichen Statistiken, wie viele SchülerInnen diesen Unterricht besuchen. In der Stichprobe einer neuen Untersuchung zu Mädchen und jungen Frauen mit griechischem, italienischem, jugoslawischem, türkischem und Aussiedlerhintergrund berichten insgesamt 63%, muttersprachlichen Ergänzungsunterricht besucht zu haben; bei den Schülerinnen und Absolventinnen mit griechischem, italienischem und türkischen Hintergrund besuchten ihn die überwiegende Mehrheit länger als 3 Jahre (BMFSFJ (Bundesministerium für Familie 2004: 246f.).

native als submersion, oder salopp „sink or swim“ bezeichnet. Hier ist nicht ausgeschlossen, dass einzelne LehrerInnen doch auf die möglicherweise mangelnden Sprachkenntnisse eingehen oder gegebenenfalls spontane Übersetzungshilfen leisten, dies geschieht aber nicht systematisch.

## 5. Ergebnisse und methodisches Vorgehen von Evaluationsstudien bilingualer Schulprogramme

### 5.1. Methodische Qualitätsmerkmale und Probleme

Bevor in Kapitel 5.2 mehrere Evaluationsstudien vorgestellt werden, die von Greene (1998) und Slavin/Cheung (2005) als methodisch akzeptabel beurteilt und in ihren Metaanalysen berücksichtigt wurden, sollen die dort angewandten methodischen Qualitätskriterien vorgestellt und diskutiert werden, da sie auch für die eigene methodenkritische Bewertung der Untersuchungen leitend sind. Diese Kriterien orientieren sich an den Merkmalen eines experimentellen Forschungsdesigns und sollen hier am Beispiel des Forschungsgegenstands der bilingualen Unterrichtung erläutert werden (vgl. zusammenfassend Kromrey 1998: 97; Fitz-Gibbon/Morris 1978: Kap. 2; Cook/Campbell 1979; Frey/Frenz 1982). Eines der wichtigsten Ziele einer solchen Experimental- bzw. Evaluationsstudie – hier verstanden im Sinne einer Wirkungsanalyse zum outcome, nicht der Implementation eines Programms (zu unterschiedlichen Evaluationstypen vgl. Kromrey 1998: 97ff.) – ist es, systematisch zu klären, ob zwischen der Maßnahme („treatment“), in diesem Fall die zweisprachige Unterrichtung, und einem Effekt eine kausale Beziehung besteht, ob also die Wirkung wirklich dem durchgeführten Programm zugerechnet werden kann oder durch andere Faktoren (mit) hervorgerufen wurde. Ein aussagekräftige Evaluation muss dafür die Frage beantworten können, mit wem und mit was die Ergebnisse, die die Maßnahme erbringt, verglichen werden (vgl. Cook/Campbell 1979: 96ff.).

Bei der Evaluation bilingualer Programmen stellen die SchülerInnen, die in dieser Form unterrichtet werden, (und seltener andere „Einheiten“ wie Schulen oder Lehrer) die **Experimentalgruppe** dar. Diese müssen mit einer **Kontrollgruppe** verglichen werden, die nicht an diesem Programm teilnimmt. Während in anderen Themengebieten diese Kontrollgruppe manchmal gar keine „Behandlung“ erfährt, gehen hier die SchülerInnen der Kontrollgruppe in den Regelunterricht oder ein alternatives, monolinguales Förderprogramm. Meist handelt es sich bei einer Evaluierung solcher Programme um Feld- im Gegensatz zu Laboruntersuchungen, weil das natürliche Umfeld, insbesondere die Konzeption und Umsetzung der zwei- bzw. einsprachigen Unterrichtung von den ForscherInnen selbst nicht maßgeblich beeinflusst wird (vgl. Bortz 1999: 8).

Aus methodischer Sicht ist eine Aufteilung der teilnehmenden Personen auf Experimental- und Kontrollgruppe nach dem Prinzip des kontrollierten Zufalls, die so genannte **Randomisierung**, ideal. Denn nur so wird ausgeschlossen, dass bezüglich einflussreicher Merkmale der Personen keine systematische Verzerrung stattfindet (z.B. dass durchschnittlich intelligentere Kinder oder mehr Kinder aus unteren Schichten am bilingualen Programm teilnehmen) und somit Unterschiede in den gemessenen Ergebnissen zwischen diesen beiden Gruppen ausschließlich auf Teilnahme/Nicht-Teilnahme an der Maßnahme zurückgeführt werden können (interne Validität) (zur Diskussion typischer Fehlerquellen vgl. Cook/Campbell 1979: 51-55; Jacobs 1998).

Wenn eine solche Randomisierung bei einer dann als Quasi-Experiment bezeichneten Studie nicht möglich ist, müssen die verbleibenden Unterschiede offen gelegt und vor allem durch die möglichst umfassende Erhebung aller weiteren potentiellen Einflussgrößen (z.B. vorherige Leistungstests als Prätests, IQ-Tests, Alter, Geschlecht, Bildungs- und sozio-ökonomischer Hintergrund der Eltern, Einreisealter etc.) statistisch verglichen und kontrolliert werden.

Wenn sich die Unterschiede zwischen Experimental- und Kontrollgruppe bei wichtigen Merkmalen als zu groß herausstellen, ist ein Vergleich der Testergebnisse nicht uneingeschränkt möglich. So stellen etwa Slavin und Cheung die Bedingung, dass Differenzen im Prätest grundsätzlich nicht mehr als eine Standardabweichung betragen sollte (Slavin/Cheung 2005: 12). Dies ist damit zu begründen, dass festgestellte Effekte nicht nur mit dem Einfluss individueller Eigenschaften, sondern auch mit der Zusammensetzung einer Gruppe zu tun haben könnten: z.B. könnte ein und dieselbe pädagogische Maßnahme bei im Durchschnitt „intelligenteren“ Gruppen besser wirken (neben dem selbstverständlichen Einfluss der individuellen kognitiven Fähigkeit auf die Leistung). Bei Vergleich von nur zwei oder wenig mehr Klassen lassen sich diese Kontexteffekte im Sinne einer Mehrebenenanalyse aber nicht statistisch kontrollieren (Vorschläge zur Berücksichtigung von Kontextmerkmalen vgl. Kapitel 6).

Neben der Möglichkeit der statistischen Kontrolle solcher Drittvariablen in multivariaten Analysemodellen bietet sich als Alternative das paarweise Matching (Parallelisierte Stichproben) an, bei dem jedem Kind aus der Experimentalgruppe eineN SchülerIn aus der Kontrollgruppe zugeordnet wird, die in einer oder mehrerer dieser für die Leistungsergebnisse bedeutenden Einflussgrößen möglichst gleich oder ähnlich ist (vgl. Bortz 1999: 140, 321; Cook/Campbell 1979: Kap. 4; zu weiteren statistischen Alternativen im Umgang mit Verzerrungen bei der Stichprobenauswahl vgl. Heckman 1979).

In dem besonders gebräuchlichen statistischen Verfahren der Kovarianzanalyse werden die statistisch korrigierten anstelle der reinen Mittelwerte der Posttests der nicht-randomisierten Vergleichsgruppe verglichen (vgl. Bortz 1999: 357-360). Seltener werden multiple Regressionsanalysen angewandt, bei denen die Zugehörigkeit zu Kontroll- und Experimentalgruppe nur eine unabhängige (Dummy-)Variable neben weiteren anderen ist. So kann man die Effekte des Treatments mit denen dieser anderen Einflussfaktoren wie dem familiären Hintergrund direkter vergleichen, was für die Frage der Steuerbarkeit des Kompetenzerwerbs durch die didaktische Praxis von Belang ist.

Im Zentrum steht der **Vergleich der Posttest-Ergebnisse** von Experimental- und Kontrollgruppe, etwa die Leistungen in einem standardisierten Sprach-, Leseverständnis- oder Mathematiktest in L2 (und je nach Forschungsinteresse auch in L1). Zugleich ist bei der Evaluation von Unterrichtsmodellen ein **Vorher-nachher-Vergleich** sowohl bei der Experimental- wie bei der Kontrollgruppe unabdingbar. Denn gerade letztere machen im Regelunterricht ja auch Lernfortschritte. Allgemein sind Prätest-Ergebnisse, die das anfängliche Kompetenzniveau widerspiegeln, ein zentraler Prädiktor, der zur statistischen Erklärung der Posttest-Ergebnisse herangezogen wird.

Zu unterscheiden bleibt des Weiteren, welche Kompetenzen als Messlatte für den Erfolg des evaluierten Programms dienen – ob man in Bezug auf die Zweitsprache den Sprachstand bzw. die Sprachentwicklung oder aber Fachwissen prüfen will, für dessen korrekte Wiedergabe Sprachkompetenz eine notwendige, aber nicht hinreichende Bedingung ist. Standardisierte **Leistungstests** in der Verkehrssprache, in denen inhaltliches Wissen abgefragt wird, machen als Prä- und Posttests erst Sinn, wenn Kinder genügend L2-

Kompetenzen erworben haben, weil andernfalls Sprachdefizite nicht von fachlichen oder kognitiven zu unterscheiden sind (vgl. Slavin 2003: 5). Um die sprachliche und kognitive Ausgangsbasis zu Beginn der Maßnahme zu berücksichtigen, können z.B. ein (vor der Einschulung mündlicher) Sprachtest in L2 und/oder ein non-verbaler Intelligenztest bzw. einer in der dominierenden Sprache durchgeführt werden.

Erhebungsinstrumente, die im Prä- und Posttest den L2-Sprachstand bzw. den Fortschritt im **Zweitspracherwerb** zwischen diesen zwei Zeitpunkten messen, unterliegen nicht solchen zeitlichen Beschränkungen, ab wann ein Test aussagekräftig ist. Beim Einsatz von normbezogenen Messinstrumenten, die die Sprachentwicklung von Muttersprachlern als Maßstab haben, ist es zwar eine Frage des Alters der zu testenden Kinder, ab wann eine sprachliche L2-Entwicklung bis hin zu einem muttersprachlichen Niveau realistisch zu erwarten ist.

Die sprachwissenschaftliche Forschung deutet darauf hin, „dass sich die Schritte bei der Aneignung des Deutschen als Erstsprache, Zweitsprache und Fremdsprache in vielen wesentlichen Aspekten ähneln. Das rechtfertigt vorerst die Anwendung prinzipiell gleicher Analysegesichtspunkte bei der Beschreibung von Deutschprofilen einsprachiger und zwei- oder mehrsprachiger Kinder und Jugendlicher.“ (Ehlich u.a. 2004: 17) Dabei ist es wahrscheinlich, dass in diesen Testverfahren GrundschülerInnen mit einer anderen Erstsprache als die Unterrichtssprache im Vergleich zu den Muttersprachlern zunächst defizitär abschneiden. Forschungsethisch ist dies unproblematisch, solange Testergebnisse der wissenschaftlichen Forschung dienen und nicht, wie häufig in der schulischen Praxis, zur Selektion führt (vgl. Ehlich u.a. 2003: 14, 18; zu kriteriumsbezogenen Verfahren in der (vor-)schulischen Förderdiagnostik vgl. Reich 2001: 916).

Allerdings ist es auch eine normative (m.E. gerechtfertigte) Entscheidung, von einem anzustrebenden muttersprachlichen Niveau z.B. gegen Ende der Primarstufe als dem eigentlichen pädagogischen Ziel auszugehen. Voraussetzung sollte sein, dass diese Sprachtests eine „Unterscheidung von Sprachentwicklungsstörung und Zweitsprachaneignung“ (Ehlich u.a. 2004: 19) zulassen. Eine zusätzliche Erhebung des Sprachstands in L1, insbesondere bei noch fehlenden L2-Kenntnissen wie etwa bei neu zugewanderten Kindern, wäre darüber hinaus ein weiterer Prädiktor für das Potenzial sprachlicher L2-Entwicklung (bzw. bei entsprechendem Forschungsinteresse für die L1-Entwicklung ohnehin geboten).

Vor dem Hintergrund dieser hier diskutierten methodischen Aspekte der Evaluationsforschung im Allgemeinen und in Bezug auf bilinguale Programme im Besonderen sind auch die methodischen Auswahlkriterien der Untersuchungen bilingualer Programme bei Greene (1998: 3) und Slavin und Cheung (2005: 11-13) zu betrachten:

- Es werden SchülerInnen eines zweisprachigen Programms mit einer Kontrollgruppe (vor allem hinsichtlich ihrer Ethnizität und ihres sprachlichen Backgrounds) vergleichbarer SchülerInnen verglichen, die nur auf Englisch unterrichtet werden.

In den beiden Metaanalysen wird nicht unterschieden, ob die Schüler in den Kontrollgruppen keinerlei extra Förderung erhalten (submersion) oder eine Form von Englisch als Zweitsprache (z.B. immersion strategies) gelehrt bekommen. Zum Teil lässt sich in den Originalstudien die Unterrichtsart in den Kontrollklassen auch nicht mehr rekonstruieren.

- Die Aufteilung der SchülerInnen auf Experimental- und Kontrollgruppe erfolgte nach dem Zufallsprinzip. War dies nicht der Fall, werden bei der nicht-randomisierten

Gruppenzuteilung Unterschiede zwischen Experimental- und nicht-äquivalenter Kontrollgruppe bei einflussreichen Variablen wie Prätests statistisch kontrolliert.

- Unterschiede in den Leistungstests zwischen Experimental- und Kontrollgruppe wurden durch angemessene statistische Tests festgestellt.

Gemäß der Prinzipien der Experimentalforschung schließen Slavin und Cheung aus methodischen Gründen außerdem Studien aus,

- wenn Prätests nicht erwähnt oder aber nach Beginn des Programms (aber zu Beginn der Evaluation) erhoben werden, und daher die Wirkung bis zu diesem Zeitpunkt ignoriert wird.

Des Weiteren wurden Studien einbezogen, für die gilt, dass die Ergebnisse, die gemessene Leistung, auf standardisierten Tests<sup>14</sup> in englischer Sprache basieren. Zusätzlich schränkt Greene (1998: 3) inhaltlich die Auswahl so ein, dass nur Studien berücksichtigt werden, die den Effekt von zweisprachigen Programmen nach mindestens einem Schuljahr messen.

Felduntersuchungen, die Unterrichtsmodelle (oder auch andere bildungs- oder arbeitsmarktpolitische Interventionen u.ä.) evaluieren, sind mit typischen methodischen Problematiken, insbesondere einer fehlenden Randomisierung, experimentellen Einbußen und einer ungenügenden Verallgemeinerbarkeit konfrontiert, auf die im Folgenden eingegangen wird:

In den Metaanalysen von Greene und Slavin/Cheung werden gemäß der Kriterien des klassischen Experiments randomisierte Forschungsdesigns als anzustrebendes Ideal dargestellt. Eine solche Randomisierung ist allerdings nicht unproblematisch (und wurde in nur wenigen, meist aus den 1970ern stammenden US-amerikanischen Studien angewandt). Bei Evaluationen zweisprachiger Schulprogramme handelt es sich bisher ausschließlich um Feld-, nicht um Laborexperimente. Auf jeden Fall ist die echte Schulzeit von Schülerinnen und Schülern betroffen. Einer Aufteilung von Schülern auf eine Experimental- und Kontrollgruppe nach dem Zufallsprinzip, also wie bei der Ziehung der Lotto-Zahlen, stehen forschungsexterne Hindernisse entgegen (Baker 2001: 230): Ohne die Zustimmung der Eltern ist eine solche Zuteilung nicht durchführbar. Bei einem gegebenenfalls mehrjährigen Unterrichtsmodell ist kaum vorstellbar, dass Eltern gänzlich gleichgültig ist, in welchen Sprachen ihr Kind unterrichtet wird (Moran/Hakuta 2001: 456). Und wenn es gewissen Eltern egal wäre, so würden diese (und ihre Kinder) sich wohl systematisch von denen unterscheiden, die einer solchen Randomisierung ihre Zustimmung verweigern. Ein Selektions-Bias bei der Auswahl der Experimentalgruppe kann also durch die fehlende Möglichkeit einer Randomisierung im Grunde nicht vermieden werden. Realistischer ist es, in Quasi-Experimenten diese Einstellungsunterschiede systematisch zu berücksichtigen, die elterlichen Präferenzen mit ihren sonstigen Eigenschaften zu verknüpfen und, wie oben (S. 20)

---

<sup>14</sup> Wie die genauen Tests aussehen, ist in den wenigsten Originalstudien und erst recht nicht in den Metaanalysen einsehbar. In den Originalstudien werden zumindest die Namen der Tests genannt, bei denen es sich meist um bereits entwickelte standardisierte Leistungstests für die gesamte USA oder einzelne Bundesstaaten handelt. Im Rahmen dieses Berichts wird davon ausgegangen, dass es sich hier um vielfach erprobte Messinstrumente handelt, die die Gütekriterien der Validität und Reliabilität erfüllen. Dass dies eine Voraussetzung ist, die bei der Neuentwicklung von Testinstrumenten keineswegs gegeben sein muss, zeigen etwa Kritiker der Berliner Sprachstandserhebung Bärenstark (Gogolin u.a. 2001b).

erläutert, alle anderen möglichen Einflussfaktoren bei der statistischen Analyse der Effekte zu kontrollieren bzw. Matching-Verfahren zu verwenden.

Als ein weiteres Problem sind mögliche experimentelle Einbußen zu nennen: Ein nicht zufallsbedingter, verzerrender Schwund in den Vergleichsgruppen (z.B. Wegzug oder Überweisung auf eine Sonderschule) kann dazu führen, dass diese in kurzer Zeit nicht mehr äquivalent sind. Diese Veränderungen wirken sich gerade bei kleinen Vergleichsgruppen, wie sie Schulklassen darstellen, statistisch gravierend aus (vgl. Thomas/Collier 1997: 21), und die Ursachen festgestellter Leistungsunterschiede und die Wirkung des Programms lassen sich dann nicht mehr unverzerrt feststellen (vgl. Slavin/Cheung 2003: 6).

Dieses Problem hat sich etwa bei Evaluierungen der in den USA am weitesten verbreiteten Early-Exit-Transitional-Bilingual-Education-Programmen gestellt. Hier gehört es zum pädagogischen Konzept, dass SchülerInnen mit guten Fortschritten in ihren Englisch-Kenntnissen das Spezialprogramm so früh wie möglich verlassen. Misst eine Forscherin zu einem späteren Zeitpunkt die Schulleistungen als Posttests, können so in der Experimentalgruppe nur die Kinder mit mäßigeren Englischkenntnissen übrig geblieben sein.

Um solche Veränderungen zu berücksichtigen, müssten die aus dem Programm ausgeschiedenen SchülerInnen weiterverfolgt werden. Alleine auf Basis von Schülerakten, auf die sich viele US-ForscherInnen stützen, ist dies nicht möglich, sondern erfordert eigene Verlaufsstudien.

Ferner ist die Verallgemeinerbarkeit vieler Evaluationsstudien oft eng begrenzt. Bei der Evaluation bilingualer Programme in einzelnen Klassen oder Schulen ist die Stichprobengröße typischerweise eher klein. Dies kann sich einerseits so auswirken, dass Effekte nicht statistisch signifikant werden (was prinzipiell pro und contra zweisprachige Erziehungsmodelle sprechen kann). Andererseits leidet darunter die Repräsentativität (externe Validität) – die Verallgemeinerbarkeit und Übertragbarkeit der Ergebnisse auf andere als den selbst untersuchten (lokalen und nationalen) Kontext oder andere als die untersuchten Sprachen (vgl. auch Thomas/Collier 1997: 21).

Schließlich sei an die prinzipiell möglichen Zielsetzungen einer Untersuchung bilingualer Programme auf ihre „L2-Wirkung“ hin erinnert: Eine Grundentscheidung in der empirischen Herangehensweise sowie in der bildungspolitischen Zielsetzung ist es, welche Frage durch eine Evaluierung beantwortet werden soll. Die klassische Frage der Wirkungsforschung lautet: Ist das (zweisprachige) Unterrichtsmodell besser, d.h. effektiver, als der Regelunterricht bzw. ein alternatives Förderprogramm? Der Vergleich findet dabei zwischen der (möglichst vergleichbaren) Subpopulation der zweisprachigen Schüler statt. Ob ein Programm besser als ein anderes wirkt, lässt aber offen, ob selbst dieser größere Leistungszuwachs ausreichend im Vergleich zu den Muttersprachlern ist. Thomas und Collier halten hier entgegen, „what really matters is how schools are able to assist English learners, as a group, to eventually match the achievement characteristics of native-English speakers, in all areas of the curriculum [...] irrespective of the students' original background“ (1997: 19f). Letztlich sollte eine Evaluation beide Fragen beantworten können.



## 5.2. Ausgewählte US-amerikanische Evaluationsstudien: Kurzdarstellung und methodische Kritik

Im Folgenden sollen einige ausgewählte Evaluationsstudien vorgestellt werden. Dies dient einerseits dem Zweck, Aufschlüsse zur Effektivität zweisprachigen Unterrichts anhand konkreter Programme zu gewinnen. Andererseits sind hier die methodischen Aspekte des Forschungsdesigns von Interesse. Deshalb werden hier drei Studien ausgewählt, die von der methodischen Qualität her zu den besseren Studien zählen:

- **Ramirez u.a. (1991a, 1991b, 1991c)**

- **Huzar (1973)**

Von ähnlicher methodischer Qualität ist auch die Studie von Plante (1978), bei Huzar ist jedoch die Stichprobe größer und das Unterrichtsmodell sieht eine gleichzeitige Alphabetisierung in L1 und L2 vor – ein für die gegenwärtige Situation in den USA untypisches, aber z.B. derzeit in den bilingualen Hamburger Schulmodellen praktiziertes zweisprachiges Unterrichtsmodell.

- **Alvarez (1975)**

Vergleichbares methodisches Niveau und ähnliche Ergebnisse weist auch die Studie von Cohen (1975) auf, deren Sample jedoch kleiner als das von Alvarez ist.

Die in den Metaanalysen berücksichtigten Evaluationsstudien sollen zwar die dort gesetzten Mindeststandards erfüllen, etliche haben sich aber bei näherer Durchsicht durch Slavin und Cheung (2005) bzw. durch die Verfasserin als fehlerhaft erwiesen:<sup>15</sup>

- Bei Covey (1973) liegt entgegen der eigenen Darstellung keine echte Randomisierung bei der Aufteilung der SchülerInnen in Experimental- und Kontrollgruppe vor, lediglich die Gesamtstichprobe wurde nach dem Zufallsprinzip gezogen (Covey 1973: 56-58).
- Bei Danoff (1978) wurden die so genannten Prätests nach Beginn des Treatments durchgeführt, so dass die Wirkung des Programms davor unberücksichtigt bleibt.
- Bei Maldonado (1994) erscheint die Stichprobe von N = 20 unvertretbar klein, zumal es sich im Gegensatz zu den anderen Studien um ein bilinguals Sonderschul-Programm handelt (vgl. zu weiteren Kritikpunkten Rossell/Kuder 2005: 66).
- Bei Powers (1978) stellt sich das Problem des Selektionsbias: Es wird in der Kovarianzanalyse nicht berücksichtigt, dass sich die Vergleichsgruppen bei der Einschulung signifikant im Sprachgebrauch unterscheiden.
- In der Studie von Saldate u.a. (1985) erscheint eine ungesteuerte Stichprobenverkleinerung zwischen Prä- und Posttest um 39% (von N = 62 auf N = 38) inakzeptabel hoch.

---

<sup>15</sup> Auch andere neue US-Evaluationsstudien, die in den bisherigen Metaanalysen nicht aufgenommen wurden, weisen solche methodischen Mängel auf, dass Kausalaussagen zur Wirkung von Programmen nicht zulässig sind: z.T. fehlt bei nicht-randomisierten Programmvergleichen die statistische Kontrolle von Prätests oder anderen zentralen Einflussgrößen (Alanís 2000; Hofstetter 2004), während etwa bei López und Tashakkori (2004) in der Untersuchung eines two-way Programms keine separaten Analysen für nicht-englische Muttersprachler aufgeführt werden. Ebenfalls nicht berücksichtigt wird die mit relativ hohen Fallzahlen operierende Studie von Lindholm-Leary (2001): Hier werden nur unterschiedliche Typen von zweisprachigen Programmen und nicht noch zusätzlich rein englischsprachige Unterrichtsmodelle verglichen und zudem keine multivariaten Verfahren verwendet.

- Bei Rossell (1990) sind Experimental- und Kontrollgruppe nicht vergleichbar. „Asiatische“ Kinder, die in den multivariaten Analysen signifikant besser abschneiden, sind nur in den monolingualen Klassen eingeschult, bzw. entspricht das englisch-chinesische Programm, an dem ca. 10 „Asiaten“ teilnehmen, nicht der Definition der anderen bilingualen Programme, weil nicht in L1 alphabetisiert wird.
- In Skoczylas (1972) Studie über ein bilinguals two-way-Programm werden nur Durchschnittswerte von Leistungstests für die gemeinsam unterrichteten englischen und spanischen Muttersprachler zusammengefasst angegeben, nicht aber separat die Englisch-Ergebnisse nur für die Kinder mit spanischer Erstsprache.

Andere Studien befassen sich mit Fragestellungen, die für die dieses Berichts inhaltlich wenig relevant sind:

- Die meisten getesteten Kinder bei Bacon u.a. (1982) sprechen Englisch als Erstsprache, zudem liegen keine Lese-Prätests vor.
- Kaufman (1968) untersucht die Wirkung von muttersprachlichem Unterricht in der Sekundarstufe. Da die Alphabetisierung in L2 schon lange vorher stattgefunden hat, bleibt theoretisch ungeklärt, warum in diesem Alter Leseunterricht in L1 überhaupt L2-Kompetenzen noch unterstützen sollte – Cummins Hypothesen greifen hier zumindest nicht.
- Rothfarb u.a. (1987) messen nur mündliche Englischkenntnisse, was für die Entwicklung der academic language weniger aussagekräftig ist.

Nur im Rahmen der Diskussion der Metaanalyse von Slavin und Cheung (2005) werden die dort angeführten Studien von Maldonado (1977) und Campeau u.a. (1975) berücksichtigt.

Während es den Rahmen dieses Berichts sprengen würde, die von Slavin und Greene aus geschlossenen Evaluationsstudien insgesamt noch einmal selbst zu begutachten und vorzustellen, so soll doch für die groß angelegten Studien von

- **Thomas und Collier (2002; 1997)**

eine Ausnahme gemacht werden, da diese ForscherInnen international erhebliche Beachtung erfahren haben.

## **Ramirez u.a. (1991a; 1991b; 1991c)**

Eine der umfangreichsten Evaluationsstudien, die von Greene (1998) und Slavin (2003) akzeptiert und z.B. auch von Reich und Roth (2002: 18) hervorgehoben wird, wurde von Ramirez und Kollegen (1991a; 1991b; 1991c) vorgelegt. Das Hauptziel dieser Studie war es, die Effektivität dreier Programmtypen zu vergleichen, die das gleiche pädagogische Ziel hatten, nämlich den Erwerb von Englischkenntnissen, so dass die SchülerInnen mit spanischsprachigem Hintergrund in einem rein englischsprachigen Regelunterricht Erfolg haben könnten (Ramirez u.a. 1991a: 1).

**Evaluierte Unterrichtsmodelle (Treatment):** Die Unterrichtsformen unterschieden sich primär im Anteil und Ausmaß von Englisch als Unterrichtssprache sowie in der Gesamt-

dauer des jeweiligen Programms (Ramirez u.a. 1991a: 2, 5; 1991b: 92; zu Programmtypen vgl. auch Übersicht auf S. 16):

- Structured English Immersion Strategy (kurz “Immersion”): Unterrichtung von Englisch als Zweitsprache (ESL) im Rahmen des Sachunterrichts; höchstens gelegentliche Übersetzung bei akuten Verständigungsschwierigkeiten; eigene Klassen („sheltered immersion“) für Englisch lernende Kinder, Überweisung in die Regelklassen nach der ersten oder zweiten Jahrgangsstufe angestrebt
- Early-Exit Transitional Bilingual Education Programs (early-exit TBE): täglich 30 bis 60 Minuten Unterrichtung in der Muttersprache (der Rest auf Englisch) mit dem Ziel der „introduction of initial reading skills“ (1991a: 2); Benutzung von L1 im sonstigen Unterricht nur bei Verständnisschwierigkeiten; Überweisung in die Regelklassen nach der ersten oder zweiten Jahrgangsstufe angestrebt
- Late-Exit Transitional Bilingual Education (late-exit TBE): mindestens 40 % des Unterrichts in L1 (auch in Fächern wie Mathematik, Sozialkunde und/oder naturwissenschaftlichem Unterricht); Überweisung in die Regelklassen nach der sechsten Jahrgangsstufe

Aus ihrer Untersuchung der praktizierten Unterrichtsformen konnten die ForscherInnen schlussfolgern, dass hinsichtlich des entscheidenden Definitionskriteriums des Sprachgebrauchs von L1 und L2 durch das Lehrpersonal die Implementierungen der Programme auch ihrer nominellen Definition entsprachen.

Allerdings werden in der Literatur verbleibende Widersprüche zwischen den formalen Programmmerkmalen und der tatsächlichen Umsetzung kritisiert: So stufen Slavin und Cheung (2005: 23) die untersuchten early-exit-Programme nicht transitorisch ein (konsekutiv erst in L1, dann in L2 alphabetisierend), sondern als ein Modell einer gleichzeitigen Alphabetisierung: Denn bereits im Kindergarten wurde zu einem Anteil von 35% des Unterrichts English language arts unterrichtet. Ferner sieht es Rossell (1992: 163) als eine unnötige Zusatzbedingung von Ramirez und Kollegen an, dass alle LehrerInnen – auch die im Immersion-Programm – prinzipiell englisch-spanisch bilingual sein sollten, was die Verallgemeinerbarkeit auf die Situation in den USA insgesamt einschränkt. Schließlich merken Ramirez u.a. (1991a: 15f.) selber kritisch an, dass es entgegen der gesetzten Programmziele weder die Immersion- noch die early-exit-TBE-Programme schafften, die Kinder nach der vorgesehen Zeit von zwei bis drei Jahren mehrheitlich in Regelklassen zu überführen – und dies obwohl die Mehrheit der SchülerInnen bis zur vierten Klasse als „fluent-English-proficient“ eingestuft wurde.

Der entscheidende Nachteil der Ramirez-Studie ist allerdings, dass aus methodischen Gründen nur der Vergleich des bilingualen early-exit-Modells mit dem englischsprachigen Förderunterricht methodisch statthaft ist.

Das Forscherteam selbst widmet sich zunächst ausführlich der Frage, welche Programme man wegen der fehlenden Möglichkeit einer Randomisierung überhaupt vergleichen darf. Da in den Schuldistrikten, in denen bilinguale late-exit-Modelle untersucht wurden, kein anderes Programm angeboten wurde, kann man diese Ergebnisse nicht mit denen der Immersion und bilingualen Early-Exit-Modellen vergleichen. Festgestellte Unterschiede in den Leistungen könnten sowohl mit dem Programm als auch dem Kontext zusammenhängen. „That is, one cannot control for district or school level differences.“ (Ramirez u.a. 1991c: 44) Dass man es hier mit systematischen Unterschieden, d.h. unterschiedlichen Ausgangslagen für den Kompetenzerwerb, zu tun hat, zeigt sich bei den Vergleichen der Schülerschaft (ebenso deren Eltern, der Qualifikation der Lehrer etc.), wie sie für eine Reihe von Merkmalen durchgeführt wurden (wenn

auch ohne Angabe von t-Tests, um die Signifikanz möglicher Unterschiede von Mittelwerten und Prozentsätzen statistisch zu bestimmen). Zum Beispiel hatten Schülereltern im late-exit-TBE-Programm mehr Lesematerial (Zeitung, Zeitschriften, Bücher) in englischer und spanischer Sprache zu Hause, halfen am häufigsten bei den Hausaufgaben, die die late-exit-TBE-Lehrer wiederum häufiger aufgaben und benoteten als in den anderen Schulmodellen (1991b: 350-355; Ramírez u.a. 1991a: 12). Auf Basis dieser Ergebnisse können nur die Immersion-Programme mit early-exit-TBE, nicht aber late-exit-TBE verglichen werden.

Methodisch am überzeugendsten sind dabei die Vergleiche von Immersion und bilingualen early-exit-Programmen, wenn diese an jeweils einer Schule angeboten wurden und Prätest-Ergebnisse vorlagen, was lediglich an drei von ursprünglich 46 Schulen der Fall war. Derselbe Vergleich, bei dem diese Programme zwar in gleichen Schuldistrikten, aber nicht den gleichen Schulen liefen, kann weniger überzeugen und liefert auch keine signifikanten Ergebnisse. Bei den bilingualen late-exit-Programmen verglichen Ramírez u.a. (1991c: 287-352) zwar Implementationsformen, die sich in Dauer und Umfang der Unterrichtung in der Muttersprache unterschieden, hinsichtlich ihrer Wirkung. Da in diesem Bericht ebenso wie in den Metaanalysen von Greene und Slavin der Vergleich „Unterricht in der Muttersprache – ja oder nein?“ im Vordergrund steht, wird im Folgenden nur auf den Vergleich von Immersion und early-exit TBE an Schulen, die beides anboten, eingegangen.

**Das Sample – Experimental- und Kontrollgruppe:** Auswahlkriterium für die teilnehmenden SchülerInnen war, dass ihre Muttersprache Spanisch war, sie bei der Einschulung in den Kindergarten/Vorschule als „limited English proficiency“ (LEP) eingestuft wurden, ab dem Kindergarten durchgehend im jeweiligen Programm in den untersuchten Schulbezirken eingeschult waren und nicht mehr als 40 Tage pro Schuljahr gefehlt hatten (1991b: 50). SchülerInnen an diesen Schulen, die beide Programme anboten, gehörten zum größten Teil der zweiten Zuwanderergeneration an: 82,5% waren in den USA geboren und die überwiegende Anzahl der Übrigen länger als zwei Jahre ansässig (1991b: 316). Ca. zwei Drittel der Eltern dieser Kinder wiesen einen niedrigen sozio-ökonomischen Status auf (Einkommen, berufliche Stellung, Empfang von staatlichen Transferleistungen) (a.a.O.: 319).

Während in der Studie insgesamt  $N = 2352$  SchülerInnen untersucht wurden, verringerte sich die Stichprobengröße bei dem hier entscheidenden Vergleich (von Immersion- und early-exit-Programmen an den drei Schulen, die beides gleichzeitig anboten und Prätest-Scores hatten) erheblich – z.B. für den Lesetest am Ende der ersten Klasse auf

$N = 153$ , davon  $n_{\text{early-exit}} = 48$  und  $n_{\text{Immersion}} = 105$ ,

und für die Analyse der Leistungssteigerung von der ersten zur dritten Klasse:

$N = 114$ , davon  $n_{\text{early-exit}} = 28$  und  $n_{\text{Immersion}} = 86$  (vgl. 1991c: 60, Tab. 13).

Im Zeitraum von 1984 bis 1988 wurden die Leistungstests verschiedener Klassenkohorten der Jahrgänge 1 bis 4 über ein bis vier Jahre hintereinander erhoben (1991b: 43-46). Aus welchen Kohorten sich das hier relevante Subsample rekrutierte, ist nicht nachvollziehbar – für die statistischen Analysen wurden die Kohorten aber ohnehin gepoolt.

Die hier relevanten drei Schulen befanden sich in Kalifornien (1991b: 41, 1991c: A-1). Die Auswahl der Schulbezirke, der Schulen und der Lehrkräfte erfolgte nicht auf Basis einer Stichprobenziehung, sondern gezielt (1991b: 48). In den Schulen bzw. ihren speziellen Pro-

grammen wurden alle SchülerInnen der hier untersuchten Jahrgänge ausgewählt, die die oben genannten Kriterien erfüllten.

**Evaluationsdesign:** Es handelt sich um ein Prätest-Posttest-Design ohne Randomisierung. In den multivariaten Analysen wurde eine Vielzahl von Drittvariablen statistisch kontrolliert. Bedingt durch die eingesetzten unterschiedlichen Leistungstests wurden Analysen für die Zeiträume Kindergarten/Vorschule bis erste Klasse („K – 1-Analyse“) sowie erste bis dritte Klasse („1 – 3-Analyse“) durchgeführt. (Die Analysen für die 3. – 6. und 1.-6. Klasse nur für die bilingualen late-exite-Programme werden aus den oben genannten Gründen nicht berücksichtigt.)

**Gemessene Leistungen – Messinstrumente:** In der K-1-Analyse wurde für die Bestimmung der Lernausgangslage zu Beginn (Herbst) des Kindergartens/der Vorschule die TOBE (Tests of Basic Experiences) für Mathematik und Sprache erhoben, wobei die Tests in der für die zum Erhebungszeitpunkt dominanten Sprache durchgeführt wurde, d.h. für die meisten Kinder auf Spanisch (1991c: 94).

„The mathematics test measures students’ mastery of fundamental mathematical concepts and the terms associated with them, as well as understanding of the quantitative relationship between objects. The TOBE language test measures vocabulary, sentence structure, verb tense, sound-symbol relationships, and letter recognition.” (1991b: 56)

Deshalb sind TOBE-Scores keine Prätests im eigentlichen Sinne, erfassen aber zumindest Teilbereiche genereller kognitiver und sprachlicher Fähigkeiten zu diesem Entwicklungszeitpunkt.

Eine differenzierte Sprachstandserhebung für die Englischkenntnisse im Kindergarten wäre wünschenswerter gewesen. Aber TOBE ist dennoch aussagekräftiger als die im Nachhinein nicht mehr überprüfbare Einstufung als LEP (limited English proficient). Eine solche Einstufung verdeckt die vorhandene Varianz in L2-Kompetenzen und kann zudem unzutreffend sein (zur Kritik an der LEP-Einstufung vgl. Rossell 1992f.; 2000: 219, 222) Zwar berichten die Autoren, dass sich durch die Einengung der Stichprobe auf Kinder mit TOBE-Prätests die Ergebnisse nicht substanziell änderten (1991c: 126, 160), dennoch ist diese Beschränkung methodisch geboten.

In den Jahrgangsstufen 1, 2 und 3 (Messzeitpunkt jeweils Frühling) wurde als Messinstrument der CTBS (Comprehensive Tests of Basic Skills, Form S) für die Tests in „English language arts“, „English reading“ und Mathematik (Testsprache Englisch) jeweils in unterschiedlichen Schwierigkeitsgraden verwendet (1991b: 55f.).

„The mathematics score is based on computation, concepts, and applications. The tests contributing to the first grade language score measure sentence structure and syntax, as well as the use of plurals, verbs, comparative adjectives, and possessive pronouns [...] The first grade reading tests measure letter sound, word recognition, and reading comprehension ...” (1991b: 57).

**Analysetechniken:** Die K-1-Analysen wurden als Regressionsanalyse dargestellt (wenn auch von den ForscherInnen selber als Kovarianzanalyse bezeichnet), mit dem Unterrichtsmodell und weiteren Einflussgrößen als unabhängige Variablen. (Die Auswahl dieser Drittvariablen basiert auf nicht publizierten Voranalysen.) Für den Vergleich „1. bis 3. Klasse“ wurden mit Hilfe eines hierarchischen Linearmodells (1991: 78f) Wachstumskurven für den Leistungszuwachs in diesem Zeitraum geschätzt. Dieses statistische Verfahren schätzt zunächst den Anfangswert in der Klasse 1 (Achsenabschnitt) (bei Kontrolle u.a. des

zuvor besuchten Unterrichtsmodells), den Zuwachs bis zur dritten Klasse (Steigung) und die Krümmung dieser Leistungskurve (durch Berücksichtigung der Werte in der zweiten Klasse). Einheiten waren hier Expanded Scale Scores (ESS) des CTBS-Tests.

„The ESS form a continuous scale that covers the range of achievement from kindergarten through high school. The ESS are designed so that a difference of one point has the same meaning regardless of the magnitude of the scores.”(Ramirez u.a. 1991c: 95) Also bedeutet z.B. eine Differenz zwischen 250 und 255 (z.B. in der 1. Klasse) genauso viel Kompetenzzuwachs wie zwischen 850 und 855 (z.B. in der 6. Klasse).

Außerdem wurde ein graphischer Vergleich der Leistungs-Wachstumskurven – Trajectory analysis of matched percentiles (TAMP) – aller drei Programmtypen mit der durchschnittlichen Leistungssteigerung einer nationalen Normpopulation (norming population) durchgeführt. Die Autoren räumen selbst etliche Nachteile dieser Analyseform ein (Ramirez u.a. 1991c: 366f.): die Verwendung unkorrigierter Mittelwerte, d.h. ohne Kontrolle relevanter anderer Einflussgrößen, und von Querschnittsdaten. Das Forscherteam hielt inkonsequenterweise an dieser Darstellungsform fest, jedoch ist eine eindeutige Bewertung von Vor- oder Nachteilen sehr kontextspezifischer Programme so nicht statthaft (vgl. Moran/Hakuta 2001: 435).

**Ergebnisse:** Die Evaluationsergebnisse werden beispielhaft anhand der K-1 und 1-3-Analysen für Lesekompetenz illustriert. Die Lesetests werden hier ausgewählt, da sich auch Slavin und Cheung auf diesen Kompetenzbereich konzentrieren. Ramirez und Kollegen berichten (ohne Angabe von Details), dass sie nach einem erschöpfenden Vergleich keine systematischen Unterschiede zwischen den Schülergruppen in den zwei Programmen an der jeweiligen Schule fanden, so dass diese Analyse einem randomisierten Experiment nahe komme (1991c: 94).

Tab. 1: Ramirez u.a. 1991: Lesekompetenz-Analyse Kindergarten (Herbst) bis erste Klasse (Frühjahr) an Schulen mit Immersion und early-exit-TBE

Prädiktor-Variablen	Regressionskoeffizient	Standardfehler	t-Statistik	p-Wert
Konstante: CTBS-Score (1. Klasse)	189,905	23,420	8,109	0,000
Programm (RK: early-exit TBE):				
Immersion	-15,259**	5,310	-2,874	0,005
„Prätest“: Summe der TOBE-Scores für Sprache und Mathematik	0,137*	0,070	1,956	0,052
Schulen (RK: Schule 21):				
Schule 12	-0,475	3,612	-0,131	0,896
Schule 14	-4,224	3,669	-1,151	0,252
Fehltage in der ersten Klasse	-0,256	0,291	-0,878	0,381
Bildung der Eltern: durchschnittl. Anzahl der Jahre in Schule und Ausbildung	-0,867	0,960	-0,903	0,368
Besuch einer preschool vor dem Kindergarten (RK: kein Besuch)	5,514	5,782	0,963	0,337
Geschlecht (RK: Jungen):				
Mädchen	-4,771	4,749	-1,005	0,317
Anzahl der Bücher im Elternhaus	3,171*	1,833	1,730	0,086
Familiensprache: Eltern sprechen mit einander auch Englisch (RK: sprechen nur Spanisch)	8,942	8,915	1,003	0,318

Quelle: vgl. Ramirez u.a. 1991c: 126, Tab. 73

Anmerkung: RK = Referenzkategorie; t-Statistik: zweiseitiger Test; N = 153, davon n<sub>early-exit</sub> = 48 und n<sub>Immersion</sub> = 105; \*\* signifikant auf dem 1%-Niveau; \* signifikant auf dem 10%-Niveau

In dieser wie auch allen übrigen Tabellen der Originalstudie wird kein  $R^2$  oder ein anderer Parameter angegeben, der die Vorhersagekraft des gesamten Modells mit seinen jeweiligen Prädiktor-Variablen quantifizieren würde. Diese Missachtung methodischer Standards ist zu bedauern, an der Aussage, ob das untersuchte Unterrichtsmodell einen statistisch bedeutsamen Einfluss hat oder nicht, ändert dies jedoch nichts.

Wie aus dieser Tabelle ersichtlich ist, wiesen diejenigen Kinder, die das bilinguale early-exit-Programm besuchten, im Frühjahr der ersten Jahrgangsstufe hoch signifikant bessere Leistungen im Lesetest auf und zwar unter Kontrolle zentraler Variablen. Die Art der Darstellung, die Form einer Regressionsanalyse – und nicht nur der Vergleich der korrigierten Mittelwerte wie in älteren Studien üblich – lässt einen direkten Vergleich der Effekte des Programms mit denen anderer Einflussfaktoren zu: In diesem Fall jedoch blieb der Zusammenhang mit den meisten anderen Faktoren nicht signifikant. Fast auf dem 5%-Niveau signifikant war nur der Einfluss des „Prätests“ (hier die Summe der TOBE-Scores für Sprache und Mathematik – lesen konnten die Kinder im Kindergarten noch nicht) als Indikator für das kognitive und sprachliche Potenzial der Kinder. Auf dem schwächeren 10%-Signifikanzniveau hatte lediglich der Indikator für das kulturelle Kapital der Eltern, die Anzahl der Bücher, einen positiven Einfluss. Dass ansonsten so zentrale Einflussgrößen wie die Bildung der Eltern hier keinen Einfluss hatten, mag daran liegen, dass innerhalb dieser eher kleinen Stichprobe die Varianz bei diesen Merkmalen gering war, d.h. die Schülerschaft relativ homogen war (für diese Teilstichprobe liegen aber keine Angaben hierzu vor). Aus diesem Teilergebnis schlussfolgerten die Autoren ähnlich der Argumentation der Befürworter muttersprachlicher Unterrichtung: „This is consistent with the primary language hypothesis that teaching students to read in their first language facilitates their learning to read in a second language.“ (a.a.O.: 126)

Der klare Vorteil von early-exit-TBE in dieser ersten Jahrgangsstufe wird allerdings dadurch relativiert, dass weder in Mathematik noch für den englischen Sprachtest ein signifikanter Unterschied zwischen den Programmen festgestellt wurde. Für beide Tests hatte nur der „Prätest“ einen hoch signifikanten Einfluss (1991c: 115, Tab. 49; 121, Tab. 61). Angesichts der leichten Vorteile des bilingualen Programms nach gut eineinhalb Jahren Beschulung ist natürlich von Interesse, wie die Entwicklung weiterging. Für die anschließende Analyse (s. Tab. 2) der Leistungszuwächse von der ersten bis zur dritten Grundschulklasse sei wiederum das Beispiel des Lesetests herausgegriffen (an denselben drei Schulen, für deren Schüler Kindergarten-Prätests vorliegen).

Bezüglich des ersten Werts, der mit diesem hierarchischen Linearmodell geschätzt wurde, d.h. der CTBS-Scores in der ersten Klasse als Ausgangswert für die Entwicklung bis zur dritten Klasse, gleicht dieser der oben dargestellten K-1-Analyse.<sup>16</sup>

---

<sup>16</sup> Im Unterschied zu K-1 waren aufgrund des größeren Längsschnitts von Klasse 1 bis 3 weniger Kinder im Subsample (s.o. S. 32f. sowie bei Ramirez u.a. 1991c: 163-165) und aufgrund fehlender Signifikanz wurden die Variablen Geschlecht und Besuch einer Vorschule aus dem Modell ausgeschlossen.) Bei den Parameterwerten fällt jedoch ins Auge, dass sich die zwei Programme nicht mehr signifikant unterscheiden (mit einer 5%-igen Fehlerwahrscheinlichkeit hat an dieser Stelle nur der TOBE-„Prätest“ einen signifikanten Einfluss.

Tab. 2: Ramirez u.a. 1991: Lesekompetenz-Analyse erste Klasse (Frühling) bis dritte Klasse (Frühjahr) an Schulen mit Immersion und early-exit-TBE

Geschätzte Werte	Prädiktor-Variablen	Regressionskoeffizient	Standardfehler	t-Statistik	p-Wert
Ausgangswert CTBS 1. Klasse	Konstante: CTBS 1. Klasse	184,834	23,714	7,794	0,000
	Programm (RK: early-exit TBE):				
	Immersion	-2,897	6,889	-0,418	0,667
	„Prätest“: Summe der TOBE-Scores für Sprache und Mathematik	0,136*	0,069	1,986	0,050
	Schulen (RK: Schule 21):				
	Schule 12	3,386	3,934	0,861	0,392
	Schule 14	-2,428	4,323	-0,562	0,576
	durchschn. Fehltag 1.-3. Klasse	-0,257	0,323	-0,797	0,428
	Bildung der Eltern: durchschn. Anzahl der Jahre in Schule und Ausbildung	0,382	0,996	0,383	0,702
Durchschn. jährliche Leistungssteigerung 1.-3. Klasse	Anzahl der Bücher im Elternhaus	3,105*	1,862	1,667	0,099
	Familiensprache: Eltern sprechen mit einander auch Englisch (RK: sprechen nur Spanisch)	8,775	9,487	0,925	0,357
	Konstante: Leistungssteigerung CTBS 1.-3. Klasse	39,446	5,883	6,705	0,000
	Programm (RK: early-exit TBE):				
	Immersion	3,419	6,867	0,498	0,620
Krümmung Leistungssteigerung 1.-3. Klasse	Schulen (RK: Schule 21):				
	Schule 12	5,657	3,749	1,509	0,134
	Schule 14	3,363	3,926	0,857	0,394
Krümmung Leistungssteigerung 1.-3. Klasse	Konstante: Krümmung	-11,047	6,638	-1,664	0,099
	Programm (RK: early-exit TBE):				
	Immersion	-19,875**	7,437	-2,672	0,009

Quelle: vgl. Ramirez u.a. 1991c: 160, Tab. 97

Anm: Einheit der Test-Scores sind Expanded Scale Scores (ESS) (Erklärung s. S. 29); RK = Referenzkategorie; t-Statistik: zweiseitiger Test; N = 114, davon  $n_{\text{early-exit}} = 28$  und  $n_{\text{Immersion}} = 86$ ; \*\* signifikant auf dem 1%-Niveau; \* signifikant auf dem 10%-Niveau

Der entscheidende Wert jedoch, den dieses Modell schätzte, ist die durchschnittliche Leistungssteigerung von der ersten zur dritten Klasse: Hier bestand kein signifikanter Unterschied mehr zwischen Schülerinnen und Schülern, die das Immersion oder early-exit-TBE-Programm besucht haben. Der Vorteil des zweisprachigen Modells für Lesetests, wie er in der K-1-Analyse vorlag, verschwand also längerfristig – oder zumindest in den erfassten Jahrgangsstufen.

Das gleiche Ergebnis wies auch die 1-3-Analyse ohne Kindergarten-„Prätests“ auf, bei der die eine zusätzliche Schule ohne Kindergarten-„Prätests“ mit aufgenommen wurde und damit die Anzahl der SchülerInnen im bilingualen Modell von  $n = 28$  auf statistisch „stabiler“  $n = 67$  anstieg (1991c: 150, Tab. 91).

Diese Hauptaussage wird auch dadurch nicht relativiert, dass die Krümmung der Leistungskurve, d.h. der Lernverlauf bei den Vergleichsprogrammen, signifikant unterschiedlich ausfiel: „The English reading subtest results show an advantage for the immersion strategy programs at second grade, but no significant difference between the programs at first or third grade.“ (Ramirez u.a. 1991c: 151) Ohne die entsprechenden Daten ist eine Projektion



unterschiedlicher Verläufe über die dritte Klasse hinaus nicht möglich. Ansonsten stellte sich bei dieser Analyse keine weitere Einflussgröße als bedeutend heraus.

Bei den Mathematik- und Sprachtests bestanden in Bezug auf den durchschnittlichen jährlichen Leistungszuwachs keine signifikanten Unterschiede zwischen Immersion und early-exit TBE und auch der Leistungsverlauf (Wert der Krümmung) unterschied sich hier nicht (1991c: 154, Tab. 93; 156, 157, Tab. 95). Hinsichtlich dritter Variablen wiesen bei beiden Subtests Mädchen einen signifikant höheren Leistungszuwachs auf, im Sprachtest schnitt außerdem die Schule 12 besser ab.

**Resümee und Kommentar:** Im Großen und Ganzen ist die Schlussfolgerung aus der Studie von Ramirez u.a., dass es keine wesentlichen Unterschiede zwischen den zwei Alternativen des rein englischsprachigen Structured-Immersion- und des zweisprachigen early-exit-TBE-Programms gab, wenn man die Leistungen bis zur 3. Schulklasse in der verfügbaren, methodisch akzeptablen Stichprobe berücksichtigt. Die AutorInnen selbst resümieren ähnlich: „In sum, after four years in their respective instructional programs, limited-English-proficient students in immersion strategy and early-exit programs (as defined in this study) demonstrate comparable skills in mathematics, language, and reading when tested in English.” (1991a: 20) Dies mag daran gelegen haben, dass sich angesichts des tatsächlichen, eher geringen zeitlichen Umfangs des spanischsprachigen Unterrichts die Programme nicht „genügend“ unterschieden (vgl. auch Rossell 1992: 165ff.). Da aber das Ziel auch bei bilingualen early-exit-Programmen ist, mittels einer Basisalphabetisierung in L1 den L2-Erwerb zu erleichtern, hätte im Sinne dieser Grundannahme ein Vorteil gegenüber dem rein englischsprachigen Immersion-Modell zu Tage treten sollen.

Da diese Studie eine der aufwendigsten und teuersten Evaluationsstudien zu bilingualen Programmen in den USA überhaupt war, wurde sie von mehreren WissenschaftlerInnen einer kritischen Beurteilung unterzogen (Baker 2001: 250; Moran/Hakuta 2001: 452-456; vgl. auch Baker 1992; Rossell 1992; Meyer/Feinberg 1992). Hinsichtlich methodischer Mängel, wird die fehlende Vergleichbarkeit der (guten) Ergebnisse der bilingualen late-exit-Modelle mit den anderen untersuchten Unterrichtsmodeilen kritisiert. Aus wissenschaftlichen wie bildungspolitischen Gründen wird bemängelt, dass einerseits keine anderen „starken“ bilingualen Programme verglichen wurden, die etwa gleichzeitige in L1 und L2 in sprachlich gemischten Klassen alphabetisieren (Two-Way Bilingual Education). Daher sei die Ramirez-Studie nicht geeignet zur Beurteilung zweisprachiger Unterrichtung an und für sich. Andererseits wurden auch weit verbreitete rein englischsprachige Förderprogramme (Englisch als Zweitsprache in Form eines reinen Sprachunterrichts in separaten Gruppen bei sonstiger Beschulung in den Regelklassen – ESL pullout taught as a subject) nicht verglichen. Zudem hätten Kinder mit einer anderen Herkunftssprache als Spanisch untersucht werden sollen. Allerdings hätten, wie es in der Ramirez-Studie deutlich wird, solche zusätzlichen Vergleiche methodisch nur Sinn gemacht, wenn die unterschiedlichen Modelle jeweils an derselben Schule durchgeführt worden wäre. Schließlich wäre wünschenswert gewesen, wenn in den multivariaten Analysen die erhobene Einschätzung der LehrerInnen zu „special learning needs“ der einzelnen Kinder wie etwa Lernbehinderungen (Ramirez u.a. 1991 b. 318) berücksichtigt worden wäre.

In der Meta-Analyse von Slavin und Cheung (2005: Tab. 1) ebenso wie in der von Greene (1998: Tab. 3) werden nur Effektgrößen (s. Erläuterung S. 55) der Lesetests in der Analyse „Kindergarten – 1. Klasse“ angegeben (bei Slavin/Cheung mit einer Effektgröße von 0,53, bei Greene 0,12 mit einem nicht-signifikanten z-Wert, weil er offenbar die Analysen mit der vierten Schule ohne Prätests zugrunde gelegt hat). Slavin und Cheung begründen ihre Entscheidung, warum sie die Analysen bis zur dritten Klasse ausgeschlossen haben, damit, dass „it uses a pretest given long after treatments had begun.“ (a.a.O.: 24). Diese Sichtweise übersieht, dass die Testwerte der ersten Klasse (dies meinen Slavin und Cheung hier mit Prätest) selber durch Variablen wie das zuvor besuchte Schulmodell und den Kindergartenprätest geschätzt werden.

Insgesamt konnte diese Evaluationsstudie keine durchgehenden, fächerübergreifenden Vorteile eines bilingualen Unterrichtsmodells mit einer Alphabetisierung in L1 zu Beginn des (Vor-) Schulbesuchs nachweisen.

## Huzar (1973)

Bei dieser Studie handelt es sich neben der von Plante (1976) um die einzige, die tatsächlich die teilnehmenden SchülerInnen per Randomisierung in bilingual unterrichtete Experimental- und Kontrollgruppe aufteilte. Aus den in Kapitel 5.1 erläuterten Gründen ist dies aus methodischer Sicht optimal (wobei aus einer forschungsethischer Perspektive fragwürdig bleibt, wie diese Aufteilung offenbar ohne Einwilligung oder Widerstand der Eltern möglich war).

**Evaluierete Unterrichtsmodelle (Treatment):** Im bilingualen Modell wurden die Kinder simultan in Englisch und Spanisch alphabetisiert: täglich 45 Minuten Sprachunterricht („reading instruction“) in Spanisch und Englisch in den Jahrgangsstufen 1 bis 3. Die Experimentalgruppe war außerdem wie folgt unterteilt: In einer Teilgruppe waren der Spanisch- und Englischlehrer bilingual, in der anderen nur der Spanisch-Lehrer. Die Kontrollgruppe erhielt täglich 45 Minuten Leseunterricht in Englisch durch eine monolinguale Lehrkraft (Huzar 1973: 34), und besuchte ansonsten offenbar den normalen Regelunterricht.

**Das Sample – Experimental- und Kontrollgruppe:** Die an der Untersuchung teilnehmenden SchülerInnen waren spanischsprachige Zweit- und Drittklässler an einer Schule, die in einem als sozial benachteiligt klassifizierten Wohnviertel der Stadt Perth Amboy, New Jersey, lag. Somit besteht die Evaluation aus zwei Teilstudien (a.a.O.: 33 und Appendix A):

Zweitklässler, die zwei Jahre bilingual unterrichtet wurden und ihre Kontrollgruppe:

$$n_{2E} = 41 \text{ und } n_{2C} = 40, \text{ somit } N_2 = 81$$

Drittklässler, die drei Jahre bilingual unterrichtet wurden, und ihre Kontrollgruppe:

$$n_{3E} = 43 \text{ und } n_{3C} = 36, \text{ somit } N_3 = 79$$

Jede dieser vier Gruppen setzte sich noch einmal aus zwei gleich großen Klassen zusammen.

Es liegen keine detaillierten Angaben dazu vor, wie genau die Gruppen aus englisch-spanisch zweisprachigen und spanisch einsprachigen Kindern zusammengesetzt (vgl.

a.a.O.: 50) war. Ebenso wenig wurden weitere Merkmale der SchülerInnen oder ihrer Eltern eigens erhoben.

**Evaluationsdesign:** Es handelt sich um ein randomisiertes Posttest-Design (a.a.O.: 39). Die Randomisierung wird wie folgt beschrieben: „When the program was initiated in 1969, the subjects were randomly assigned to either bilingual or regular classes, which remained intact throughout the primary grades.” (a.a.O.: 34) Dies geschah aber nicht durch die Forscherin selbst, sondern durch die Schulverwaltung.

Um die Vergleichbarkeit von Experimental- und Kontrollgruppen zu überprüfen, wurden diese zudem anhand folgender zwei Tests verglichen:

**Gemessene Leistungen – Messinstrumente:** Als Ersatz eines Prätests dienten der Metropolitan Readiness Test (METRO), der bei der Einschulung die Schulfähigkeit messen soll.

Er besteht aus einer Reihe von Teilttests (z.B. Bedeutung von Wörtern, Erkennen von Buchstaben, visuelle Wahrnehmung). Da METRO nur auf Englisch durchgeführt wurde, wurden damit bei Kindern aus spanischsprachigen Elternhäusern sicher nicht nur das kognitive Entwicklungsstadium oder Intelligenz an sich, sondern auch die englischen Sprachkenntnisse gemessen.

Bei Drittklässlern wird zudem der Lorge-Thorndike-Intelligenztest durchgeführt. Als Posttest für Leseverständnis diente die Inter-American Series in englischer und spanischer Sprache (a.a.O.: 35f.).

**Analysetechniken:** Bei einem Vergleich der Posttest-Mittelwerte von Experimental- und Kontrollgruppe wurden t-Tests durchgeführt.

**Ergebnisse:** Nachdem – ohne Angabe der exakten t-Tests – festgestellt wurde, dass sich Kontroll- und Experimentalgruppe nicht signifikant im METRO- und dem Intelligenztest unterschieden, wurde die Lesekompetenz untersucht (a.a.O.: 39):

Tab. 3: Huzar 1973: Lesetest-Vergleich bilingual und monolingual alphabetisierter SchülerInnen

Klasse	Programm	Mittelwert	Varianz	Standardfehler	Differenz Mittelwert	t-Statistik Huzar	n
2. Kl.	bilingual	35,51	10,89	1,70	0,16	0,628	41
	monolingual	35,35	12,33	1,95			40
3. Kl.	bilingual	61,16	21,00	3,20	5,1	1,3844	43
	monolingual	56,06	7,47	1,25			36

Quelle: vgl. Huzar 1973: 42f; 44f., Tab. 1 und 2

Nach Angaben von Huzar waren die Unterschiede in den Mittelwerten der Lesetests weder bei den Zweitklässlern, noch den Drittklässlern auf dem 5%-Niveau signifikant verschieden. Auf Basis welcher Gleichung Huzar zu diesen t-Statistiken kam, ist nicht nachvollziehbar. Denn die Mittelwertdifferenz zwischen den Vergleichsgruppen ist bei den Drittklässlern mit 5,1 wesentlich höher als in der zweiten Jahrgangsstufe mit 0,16. (Nur über den Zugang zu den Individualdaten wäre es möglich nachzurechnen, ob bei ersteren die Testmittelwerte nicht doch signifikant unterschiedlich sind).

Ins Auge fällt, dass bei den Drittklässlern die Varianz der Leistungstests von bilingual unterrichteten Kindern beinahe dreimal so hoch war wie die der Kontrollgruppe. Huzar selbst

versuchte nicht zu erklären, warum dies sein könnte. Diese höhere Varianz ebenso wie die höhere durchschnittliche Lesekompetenz wird folgenden Grund gehabt haben: Die bilingual unterrichteten Drittklässler, deren Englischlehrer auch bilingual war, schnitten signifikant besser ab als die bilingual unterrichteten SchülerInnen, deren Englischlehrer einsprachig war (a.a.O.: 49, Tab. 5). Ob dieser Unterschied wirklich am Sprachvermögen dieser Lehrer oder an anderen Eigenschaften lag, kann auch nach Ansicht von Huzar (a.a.O.: 51) in dieser Untersuchung nicht geklärt werden. Zumindest hatte dieser Unterschied nicht mit der zusätzlichen Unterrichtung der Muttersprache an und für sich zu tun.

Ferner ergab sich bei einer nach Geschlecht – nicht aber Jahrgangsstufen – differenzierten Analyse, dass bilingual unterrichtete Jungen signifikant besser abschnitten, während bei den Mädchen kein Unterschied bestand.

**Resümee und Kommentar:** Bei der Untersuchung zweier Jahrgangsstufen hatte für die Jüngeren die zusätzliche Alphabetisierung in der Muttersprache keine Auswirkung auf die Leseleistung im Englischen. Bei Drittklässlern scheint ein Vorsprung der bilingual unterrichteten Kinder weniger mit dem Spanischunterricht an sich, als einer Variante des Treatments (mono- versus bilingualer Englischlehrer) zu tun gehabt zu haben, ohne dass hier eine Kausalwirkung endgültig geklärt werden könnte. Insgesamt beruht die gesamte Analyse darauf, dass eine nicht von der Wissenschaftlerin selbst durchgeführte Randomisierung korrekt vonstatten ging und so zur Feststellung eines möglichen Kausaleffekts die statistische Kontrolle anderer Einflussgrößen nicht notwendig war. Gerade die Unterschiede zwischen Mädchen und Jungen deuten aber darauf hin, dass hier ein multivariates Analyseverfahren verlässlichere Ergebnisse geliefert hätte. Insbesondere wäre interessant gewesen, unterscheiden zu können, wie SchülerInnen, die bei ihrer Einschulung bereits bilingual waren oder noch gar keine Englischkenntnisse besaßen, in den zwei Programmen abschnitten.

Greene (1998: Tab. 3) gibt in seiner Metaanalyse – unter Angaben nicht ganz korrekter Stichprobengrößen für die Drittklässler – eine Effektgröße an, die ungerechtfertigterweise für beide Kohorten zusammen errechnet wurde. Slavin und Cheung (2005: 22) führen für Zweitklässler und Drittklässler niedrige Effektgrößen (0,01 bzw. 0,31) an, wobei sie die signifikante Wirkung bei Jungen im Gegensatz zu Mädchen hervorheben.

Insgesamt ist aus der Evaluation von Huzar zu schließen, dass die hier untersuchte Form der gleichzeitigen Alphabetisierung in L1 und L2 keine eigenständige Wirkung auf die L2-Lesekompetenz hatte.

## **Alvarez (1975)**

In der Evaluationsstudie von Alvarez wurden zwei Implementationsvarianten bilingualer Modelle evaluiert, wie sie Ende der 1960er, Anfang der 1970er Jahre durch das damalige US-weite Förderprogramm Bilingual Education – Titel VII der Elementary and Secondary Education Amendments von 1967 – gefördert wurden.

**Evaluierte Unterrichtsmodelle (Treatment):** Da in dieser historischen Phase bilinguale Programme erst entwickelt wurden, wurde das hier praktizierte Unterrichtsmodell vom Autor selbst nicht klar einer der inzwischen etablierten Programmbezeichnungen zugeordnet, wie sie in Kapitel 4 vorgestellt wurden. Insgesamt ähnelt es einem two-way late-exit-

Transitional-Bilingual-Education-Programm (bzw. Maintenance/Developmental Bilingual Education): Das gesetzte Ziel war eine altersgerechte Sprachkompetenz in L1 und L2 (a.a.O.: 85). Eine Alphabetisierung erfolgte zunächst in L1 – für die spanischen Muttersprachler also Spanisch. Erst bei ausreichenden mündlichen L2-Kenntnissen begann das Lesen- und Schreibenlernen in L2 (a.a.O.: 58).<sup>17</sup> Vom Umfang der Unterrichtung in L1 und L2 her handelt es sich dabei um ein 50:50-Programm. In der hier untersuchten zweiten Jahrgangsstufe wurde der Sachunterricht in Mathematik, natur- und sozialwissenschaftlichen Fächern jeweils in der Muttersprache getrennt, Fächer wie Sport, Musik und Kunst abwechselnd von englisch- und spanischsprachigen Lehrern unterrichtet (a.a.O.: 73f). In dieser Implementationsphase war unklar, ob sich aus den Angeboten ein late- oder early-exit-Programm entwickeln würde – nur eine der zwei untersuchten Schulen bot zweisprachige Klassen bis zur fünften Klasse an (a.a.O.: 60). Bis zur hier untersuchten zweiten Jahrgangsstufe ähnelten sich aber die bilingualen Programme der zwei Schulen.

**Das Sample – Experimental- und Kontrollgruppe:** Insgesamt nahmen  $N = 147$  mexikanisch-amerikanische SchülerInnen der zweiten Klasse (1973/74) teil, die an zwei Grundschulen in Austin, Texas, in vier spanisch-englische und drei rein englischsprachige Klassen gingen. Es ergaben sich die folgenden Teilstichproben:

Schule B:  $n_B = 77$ , darunter in der Experimentalgruppe  $n_{B\ bi} = 51$ , und in der Kontrollgruppe  $n_{B\ mono} = 26$

Schule P:  $n_P = 70$ , darunter  $n_{P\ bi} = 39$  und  $n_{P\ mono} = 31$  (vgl. a.a.O.: 21, Tab 1)

Das zahlenmäßige Geschlechterverhältnis der 7-bis-8-jährigen Kinder war ausgeglichen, die meisten hatten einen niedrigen sozio-ökonomischen Status und waren „culturally and linguistically different than the dominant group“ (a.a.O.: 56).

Zu diesem Zeitpunkt war die Teilnahme am bilingualen Programm noch freiwillig und nicht wie kurz darauf für als limited-English-proficient eingestufte Kinder verpflichtend (a.a.O.:59). Nicht erläutert wird, ob dementsprechend der Besuch des bilingualen Programms nur auf Wunsch der Eltern erfolgte.

Ein Mangel der Studie ist, dass die Dauer des Treatments nicht für alle Kinder in der Experimentalgruppe gleich war: 82% wurde ab der ersten Klasse, 18 % aber schon ab dem vorangegangenen Kindergarten-Jahr bilingual unterrichtet (a.a.O.: 10). Trotz des starken Übergewichts der erst genannten Gruppe hätte dieser Unterschied in multivariaten Analysen statistisch berücksichtigt werden müssen.

Der Unterricht, den die Kontrollgruppe besuchte, wird in keiner Weise beschrieben, außer dass die Unterrichtssprache nur Englisch war. Angesichts dieser fehlenden Beschreibung des Non-Treatments lässt sich lediglich vermuten, dass es reine Regelklassen ohne gesonderte Förderung für Englisch-Lernende waren.

**Evaluationsdesign:** Es handelt sich um ein nicht-randomisiertes Prätest-Posttest-Design. Ähnlich wie schon bei der Ramirez-Studie dient der verwendete „Prätest“ der Kontrolle

---

<sup>17</sup> Wenn man nicht als Definitionskriterium strikt das spätere Einsetzen der Alphabetisierung in L2 benutzt, könnte man mit Slavin und Cheung (2005: Tab. 1) auch von einem Modell paralleler L1-L2-Alphabetisieren sprechen, für das sie den Terminus „paired bilingual model“ einführen.

vergleichbarer Ausgangslagen zum Beginn der Beschulung, auch wenn mit den anderen Testaufgaben, der anderen Skalierung etc. keine direkte Entwicklung zwischen den beiden Messzeitpunkten bestimmbar ist.

**Gemessene Leistungen – Messinstrumente:** Als Prätest wurde die Form A des Metropolitan Readiness Tests (METRO) (s. oben S. 40) am Anfang der 1. Klasse abgenommen. Mitte der zweiten Klasse wurde mit dem hier als Posttest verwendeten California Achievement Test, Level Two (CAT), Indikatoren für Lese- (Vokabeln und Textverständnis) und Mathematikkompetenzen (Konzepte und Problemlösung sowie Rechnen) in englischer Sprache erhoben (a.a.O.: 89-91).

**Analysetechniken:** Es wurden Kovarianzanalysen durchgeführt, wobei die Ergebnisse des METRO-Tests als einzige weitere Kontrollgröße als Ausgleich für die fehlende Randomisierung verwendet wurde (a.a.O.: 95f.). Für Unterschiede bei den METRO-Scores oder bei anderen Merkmalen der Kinder werden keine Signifikanz-Tests angegeben.

**Ergebnisse:** Für den Teilttest, der Leseverständnis misst, wird der Unterschied zwischen Experimental- und Kontrollgruppe für je beide Schulen B und P wiedergegeben. Entscheidend sind die auf Basis der „Prätests“ korrigierten Mittelwerte (denn anhand der Prätest-Mittelwerte ist ersichtlich, dass tendenziell die leistungstärkeren spanischsprachigen Kinder bilinguale Programme gewählt haben). Sowohl bei diesem als auch bei den drei weiteren Teilttests (s.o.) sind die Unterschiede zwischen zweisprachig und monolingual unterrichteten Hispanics, jeweils an beiden untersuchten Schulen, statistisch nicht signifikant verschieden (a.a.O.: 98).

Tab. 4: Alvarez 1975: Kovarianzanalyse für Leseverständnistest im mono- und bilingualen Unterricht: Intra-Schulvergleich

Schule	Programm	Mittelwerte des Prätests	unkorrigierte Mittelwerte des Posttests	korrigierte Mittelwerte des Posttests	F-Statistik	p-Wert	DF	n
Schule B	biling.	46,21	2,01	1,94	1,14	0,29	1; 67	55
	monoling.	38,16	2,04	2,13				26
Schule P	biling.	30,27	1,71	1,69	0,61	0,44	1; 74	39
	monoling.	26,00	1,82	1,87				31

Quelle: vgl. Alvarez 1975: 97, Tab. 3

Anmerkung: DF = Freiheitsgrade – diese Angabe zu den Freiheitsgraden, bei denen jeweils zwei Zahlen angegeben werden, konnte von der Verfasserin nicht nachvollzogen werden

Unter der Voraussetzung, dass selbstverständlich auch andere Einflüsse auf die Schulleistungen der untersuchten SchülerInnen wirkten, interpretiert Alvarez die Ergebnisse wie folgt: „...if bilingual education is accepted as the main difference in the students' experiences in English reading and mathematics, then it may be suggested that emphasis and time devoted to Spanish language appears to have produced no effects on student learning in English reading and mathematics" (a.a.O.: 101). In diesen Fächern haben bilingual unterrichtete Kinder also weder Vor- noch Nachteile, aber ihre Spanischkenntnisse sind – wenig

überraschend – signifikant besser als bei den Kindern ohne muttersprachlichen Unterricht (a.a.O.: 115).

**Resümee und Kommentar:** Grundsätzlich ist den Schlussfolgerungen des Forschers zu folgen, dass es in den in englischer Sprache gemessenen Lese- und Mathematik-Leistungstests keine Unterschiede zwischen den zwei Programmen gab. Aus methodischer Sicht bleiben Zweifel bestehen, ob mit dem Einschluss der METRO-Ergebnisse als Prätest der Korrektur der fehlenden Randomisierung genüge geleistet wurde. Auch wenn ein Prätest, der das kognitive und sprachliche Ausgangsniveau misst und dabei schon Einflüsse der eigenen Biographie und des Elternhauses wie die elterliche Bildungsnähe etc. reflektiert, sicherlich die wichtigste Kontrollvariable ist, wäre die Berücksichtigung weiterer, davon eventuell unabhängig wirkender Variablen (insb. Dauer des bilingualen Programms, Entwicklungsniveau in der Muttersprache oder weniger sprachabhängige Intelligenztests bei der Einschulung) wünschenswert. Allgemein mangelt es an einer detaillierteren und nachvollziehbareren Dokumentation der statistischen Analysen.

In Slavin und Cheungs (2005: Tab. 1) Meta-Analyse werden für beide Schulen zusammen genommen für Vokabel- und Textverständnistests nicht signifikante Differenzen zwischen Experimental- und Kontrollgruppen (Effektgrößen = +0,12 bzw. -0,23) angegeben.

Insgesamt war die Wirkung bi- und monolingualer Alphabetisierung in den von Alvarez evaluierten Unterrichtsmodellen vergleichbar.

### **Thomas und Collier (1997 und 2002)**

Im Folgenden sollen die Studien von Thomas und Collier (1997; 2002) vorgestellt werden. Diese Untersuchung dient verbreitet als Beleg für die förderlicher Wirkung der Alphabetisierung und Unterrichtung in der Erstsprache und wird z.B. in dem viel beachteten Gutachten von Reich und Roth (2002: 19f) prominent vorgestellt. In einem Bericht zur Mehrsprachigkeit von Schülerinnen und Schüler der schwedischen nationalen Bildungsagentur (Skolverket 2002: 11) wird sie als *die* beweiskräftige (nicht-schwedische) Studie überhaupt zitiert, die in das Plädoyer für mehr muttersprachlichen Unterricht einfließt. Bei Slavin und Cheung (2005) dagegen wurde sie aufgrund methodischer Mängel aus ihrer Metaanalyse ausgeschlossen. Aufgrund dieser gegensätzlichen Rezeption wird daher die Forschung von Thomas und Collier nun näher auf ihre Aussagekraft hin untersucht.

Ein Punkt, in dem sich Thomas und Collier nach eigenen Angaben von den sonst üblichen Evaluationsstudien unterscheiden, ist die Größe ihres Samples. In ihrer Kritik an den methodologischen Prämissen und Charakteristika der Experimentalforschung (vgl. auch Kap. 5.1) heben sie u.a. die mangelnde Repräsentativität der meisten Forschungsarbeiten in diesem Themenfeld hervor (Thomas/Collier 1997: 21). – Angesichts der kleinen Fallzahl auch der hier vorgestellten Studien ist dies eine berechtigte Kritik. – Sie selber präsentieren dagegen große Stichproben. Allerdings bleibt in der älteren Publikation (Thomas/Collier 1997) nicht nachvollziehbar, wie deren Sample erhoben wurde und wie es sich zusammensetzt.

In ihrer Veröffentlichung aus dem Jahr 1997 berichten die AutorInnen, dass sie 700.000 Schülerakten von „language minority students“ gesichtet hätten, von denen

42.317 SchülerInnen die untersuchten Schulen mindestens vier Jahre lang besuchten, wodurch Längsschnittanalysen möglich waren: „This represents the largest database collected and analyzed in the field of language minority education, to date“ (Thomas/Collier 1997: 30). Wissenschaftlich inakzeptabel ist, dass an keiner Stelle diese große Fallzahl nach der Art des mono- oder bilingualen Schulprogramms und irgendwelchen Personenmerkmalen aufgeschlüsselt wird. Die festgestellten „Patterns of K-12 [Kindergarten bis 12. Klasse] English learners' long-term achievement [...] compared across six program models“ (a.a.O.: 53, Figure 5 als graphische Darstellung dieses Hauptergebnisses) bleiben somit in keiner Weise nachvollziehbar und überprüfbar (und damit auch nicht die dort postulierte klare Überlegenheit von bilingualen Programmen, insbesondere solcher mit langjähriger zweisprachiger Unterrichtung und von two-way-Programmen). Zudem handelt es sich offenbar um Analysen, bei denen Testergebnisse weder durch Prätests oder andere Variablen statistisch kontrolliert wurden (zu weiteren Kritikpunkten vgl. Baker 2001: 261).

In der neueren Studie (Thomas/Collier 2002) zerfällt die Gesamtstichprobe bei näherer Analyse in kaum vergleichbare Samples aus vier Teiluntersuchungen, aus denen diese Publikation besteht, so dass von der versprochenen Repräsentativität im Grunde nichts mehr übrig bleibt. Hier sind die vier Teilstudien immerhin so genau dokumentiert, dass man Defizite und fehlende Vergleichbarkeit im Detail aufdecken kann. Bei allen Vier gehen die ForscherInnen bei der Datengewinnung gleich vor (s.u.), aber sowohl die Merkmale der Schülerschaft als auch der Schulkontext unterscheiden sich erheblich. Für die spezifische Fragestellung dieses Berichts sind drei Teilstudien nur von bedingtem Interesse.

So standen in einer Teiluntersuchung zu zwei Schulbezirken im Nordosten der USA (a.a.O.: 50ff.) Schülerinnen und Schüler frankophoner Herkunft im Zentrum, deren Erstsprache aber Englisch war. Bei der Teiluntersuchung im Nordwesten der USA (innerstädtische Schule im Bundesstaat Oregon) (a.a.O.: 207ff.) sind die Fallzahlen für die schriftlichen Leistungstests spanischsprachiger Schülerinnen und Schüler sehr niedrig (für 3. Klasse  $n = 31$ , für 5. Klasse  $n = 17$ ). In der Teiluntersuchung in Florida (a.a.O.: 258ff.) standen noch keine Daten aus standardisierten Leistungstests von Nicht-Muttersprachlern an bilingualen Schulen zur Verfügung, stattdessen wurden die Leistungszuwächse von SchülerInnen nicht-englischer Muttersprache in English-as-a-second-language-Klassen untersucht. Sollte die große Stichprobe von Thomas' und Colliers 1997er-Publikation aus ebenso divergierenden Teiluntersuchungen zusammengesetzt sein, wäre dies umso weniger akzeptabel.

Ausführlicher soll hier nur eine Teiluntersuchung (Texas) vorgestellt werden (a.a.O.: 116ff.), weil nur diese den für diesen Bericht relevanten Vergleich bi- und monolingual unterrichteter Kinder mit thematisiert bzw. entsprechende Daten bereithält.

**Evaluierte Unterrichtsmodelle (Treatment):** Untersucht wurden folgende Unterrichtsmodelle (a.a.O.: 124f.), die sich stark an den Bezeichnungen und Programmbeschreibungen anlehnen, die in Kapitel 4 beschrieben wurden. Alle drei bilingualen Programmtypen ähnelten sich in den Jahrgangsstufen Kindergarten bis dritte Klasse, in denen das Verhältnis der Unterrichtssprachen 90:10 (Span. – Engl.) betrug.

- two-way bilingual immersion für Kinder mit spanischer und englischer Muttersprache, ab der 4. Klasse Unterrichtssprachen im Verhältnis 50:50
- one-way developmental bilingual education (DBE) nur für spanischsprachige Kinder, ab der 4. Klasse Unterrichtssprachen im Verhältnis 50:50
- early-exit transitional bilingual education (early-exit TBE), ab der vierten Klasse Unterricht ganz auf Englisch

Zwei monolinguale Unterrichtsmodelle standen zum Vergleich:



- ESL-content-Programm (auch Sachunterricht mit Einsatz von Englisch-als-Zweitsprache-Methoden) [hierfür liegen nur begrenzte Ergebnisse vor]
- reiner Regelunterricht ohne besondere Förderung – auf Wunsch der Eltern in keinem der bi- oder monolingualen Förderprogramme (submersion/waivered LEP students)

**Das Sample – Experimental- und Kontrollgruppe:** Thomas und Collier stützten sich auf zwei Datenquellen:

Einerseits verwendeten sie amtliche Schulstatistiken zu Leistungstests aller SchülerInnen des Schulbezirks Houston, wobei es sich um eine Vollerhebung von Querschnittsdaten handelte (a.a.O.: 127). Für das Jahr 1999 sind Fallzahlen für Kinder der Jahrgangsstufen 2 bis 11 angegeben, wobei hier bei den nicht-englischsprachigen Schülern noch nicht die Einschränkung auf die Spanischsprachigen erfolgte (a.a.O.: 145, Figure C-1; Fallzahlen s. Tab. 5 unten):

Andererseits wurden vom Forscherteam selbst Schülerakten an Schulen, die zweisprachige Programme anboten, ausgewertet: Es handelt sich um SchülerInnen, die im den Jahren 1996-2000 für ein bis fünf Jahre in einem der genannten drei bilingualen Unterrichtstypen unterrichtet wurden (a.a.O.: 131; Fallzahlen s. Tab. 5 unten). Für diese Gruppe lagen auch „Quasi-Längsschnittdaten“, d.h. Informationen darüber vor, wie viele Jahre die Erst- bis Viertklässler jeweils bis zum Messzeitpunkt ein solches bilinguales Programm besucht hatten. Alle teilnehmenden Kinder (außer die englischen Muttersprachler) wurden bei ihrer Einschulung als limited English proficient (LEP) eingestuft.

Der Hauptaugenmerk der Autoren liegt auf dem Vergleich zwischen den unterschiedlichen bilingualen Programmen: two-way-Programme und one-way DBE jeweils im Vergleich mit early-exit-TBE-Programmen. Dabei waren diese TBE-Vergleichsgruppen in denselben Nachbarschaften angesiedelt und hinsichtlich ihres Anteils von LEP-SchülerInnen und dem sozio-ökonomischen Status der Schülerschaft vergleichbar, „matching schools that had comparable proportions of LEP students in the two categories of socioeconomic status (free & reduced lunch versus paid lunch“ (a.a.O.: 133). Genaue Prozentangaben werden diesbezüglich nicht gemacht; bei two-way-Programmen und ihrer TBE-Vergleichsgruppe seien die Nachbarschaften „somewhat more middle class“ (a.a.O.: 138) und werden daher in der Tab. 5 „Mittelschichtsviertel“ genannt.

Separat werden Ergebnisse der Schüleraktenanalyse für die Gruppe der „former LEP students whose parents refused Bilingual/ESL services“ (Jahrgangsstufen 1 bis 6) angeführt und zwar für solche als „waivered students“ bezeichneten Kinder, an deren Schulen prinzipiell ein bilinguales Programm vorhanden gewesen wäre, aber von den Eltern bewusst abgelehnt wurde (a.a.O.: 203-205). (Ob diese Schulen exakt die gleichen waren wie die des Vergleichs zwischen bilingualen Programmen, ist zu vermuten, aber nicht nachvollziehbar dokumentiert.) Eine echte Kontrollgruppe stellen diese SchülerInnen nicht dar, da trotz fehlender Randomisierung keine wichtigen Hintergrundmerkmale statistisch verglichen wurden. Lediglich – bzw. immerhin – wohnen sie und die bilingual unterrichteten Kinder in den gleichen Nachbarschaften und gehen in die gleichen Schulen, wenn auch nicht in dieselben Klassen.

Als Begründung dieser Auswahl führen die AutorInnen selbst an: “Our assumption we were testing was that perhaps the waivered students would do better in school than the district-wide analyses showed when they attended the same schools where their Hispanic counterparts were attending bilingual classes. This might be true because they would interact in social situations in the school, during recess and lunch and specials (art, music, physical education, etc.) and support each other on the social/emotional side of learning, making it a little less oppressive for these LEP students attending all-English mainstream classes from the first day of their arrival.” (Thomas/Collier 2002: 138) Die hier durchklingende Voreingenommenheit – dass monolingualer Unterricht schädlich ist – wird auch in ihrer anschließenden Interpretation der Analysen deutlich werden.

**Evaluationsdesign:** Es handelt sich um eine einfache deskriptive Kohortenanalyse. Die Ergebnisse standardisierter Leistungstests – die „Posttests“ – liegen für alle SchülerInnen für die Jahrgangsstufen 1 bis 11 in Form amtlicher Daten vor; für die Klassen 1 bis 5 stammen sie aus den selbst ausgewerteten Schülerakten. Es fehlt jedoch ein eigentlicher Prätest, der die sprachlichen und kognitiven Leistungsvoraussetzungen zu Beginn der Beschulung misst (vgl. auch Slavin/Cheung 2005: 10; Rossell 1998: 17f.). Die Einstufung als „limited English proficiency“ (LEP) zur Einschulung ist keinesfalls als Ersatz für einen wirklichen Prätest zu sehen, der eine Vergleichbarkeit trotz fehlender Randomisierung sicherstellen könnte.

**Gemessene Leistungen – Messinstrumente:** Aus den Schülerakten wurde die dort verzeichneten Testscores des „Stanford 9“ Lese-, Sprach- und Mathematik-Tests verwendet. Als Messeinheit dienten Normal Curve Equivalents (NCE), also keine absoluten oder standardisierten Werte. NCEs geben die relative Stellung in einem Leistungsspektrum pro Jahrgangsstufe an.

Bei NCEs handelt es sich um Perzentile, die statistisch so umgewandelt wurden, dass sie Werte von 1-99 annehmen, einen Mittelwert bzw. Median von 50 und ein NCE jeweils den gleichen Leistungszuwachs bedeutet; eine Standardabweichung beträgt ca. 21,06 NCE (Thomas/Collier 1997: 85). Um dieselbe relative Position im darauf folgenden Schuljahr beizubehalten, muss der Zuwachs 0 NCEs betragen, ein positiver Wert bedeutet ein Aufholen, ein negativer ein Abfallen innerhalb des Leistungsspektrums einer Referenzgruppe (norming population).

In der Studie liegen keine Informationen vor, auf welche Grundgesamtheit dieses Leistungsspektrum basierte, d.h. ob sich die angegebenen NCE-Mittelwerte für die hier untersuchten Gruppen sich auf nationale Daten oder die Daten des Staates Texas oder des Schulbezirks Houston beziehen. Prinzipiell wären Angaben hierzu geboten. Dennoch bleibt die Interpretation der Testergebnisse insofern gültig, als ein niedriger NCE-Wert eine eben „relativ“ schlechte Leistung widerspiegelt.

**Analysetechniken:** Es erfolgen nur die Angaben von bivariaten Ergebnissen (Durchschnittswerte in Kreuztabellen) ohne Angabe von Signifikanzniveaus für Unterschiede zwischen verschiedenen Schülergruppen.

**Ergebnisse:** Obwohl sich die vorgestellten Ergebnisse nicht zur Erklärung der kausalen Wirkung der Programme eignen, soll in einer eingehenden Analyse dargelegt werden, dass die AutorInnen diese deskriptiven Ergebnisse irreführend darstellen. In der Tabelle 5 werden Ergebnisse aus drei Tabellen zusammengeführt, die bei den Autoren selber getrennt voneinander aufgeführt werden.

Tab. 5: Thomas und Collier 2002: Englischsprachige Lesetests (Mittelwerte in NCE-Scores des Stanford 9-Tests), Querschnittsdaten für diverse Jahrgänge

	Schulstatistik				Thomas/Collier-Stichprobe						
Klasse	Engl. Muttersprachler im Regelunterricht	(ehemalige) limited English proficient SchülerInnen nach Unterrichtsart									Werte
		early exit TBE	Englisch als Zweitsprache (ESL)	Regelunterricht ohne Förderung	Regelunterricht ohne Förderung (an TBE-Schulen)	early exit TBE	Regelunterricht ohne Förderung (an DBE-Schulen)	one-way dev. bilingual Education (DBE)	Regelunterricht ohne Förderung (in Mittelschichtsviertel)	early-exit TBE (in Mittelschichtsviertel)	
1	-				48,7	51,5	46,5	50,1	51,8	51,9	MW n SD
					166	1299	144	1354	48	421	
					17,8	18,6	15,6	18,8	20,3	18,5	
2	49			49	49,8	47,3	43,8	47,9	40,3	44,7	MW n SD
					135	1191	168	1295	31	394	
					17,9	15,9	13,3	15,9	14,9	17,15	
3	50			45	45,0	46,8	44,3	46,9	50,2	48,9	MW n SD
					191	1321	191	1358	38	444	
					18,3	17,7	17,0	18,6	18,7	18,8	
4	49	51	57	34	46,1	44,7	40,2	44,0	46,7	44,4	MW n SD
					105	1429	184	1385	36	512	
					19,5	17,9	16,34	18,9	18,1	20,1	
5	50	53	59	32	30,3	41,4	(39,1)	41,4	36,0	38,9	MW n SD
					99	1692	17	1930	30	387	
					17,1	18,3	20,1	19,8	17,5	19,3	
6	48	54	55	29	(25,6)		(23,80)				MW n SD
					17		7				
					15,6		16,1				
7	46	48	48	26	-		-				MW
8	47	47	47	28	-		-				MW
9	45	46	43	24	-		-				MW
10	47	45	39	22	-		-				MW
11	49	47	40	25	-		-	-			MW
Fallzahl (Spalten)	103887	3333	3655	1599	713	7156	886	7391	138	2158	n
Spalten-Nr.	1	2	3	4	5	6	7	8	9	10	11

Quellen: Spalten 1-4 Thomas-Collier 2002: 155, Tab. C-1; Spalte 5, 7 und 10 a.a.O.: 204f., Tab. C-10; Spalte 6 und 8 (a.a.O.: 190-192, Tab. C-7)

Anmerkungen: Mit Ausnahme der englischen Muttersprachler wurden alle SchülerInnen zur Einschulung als limited English proficient (LEP) eingestuft. Je nach Jahrgangsstufe und Programm sind die Kinder noch oder nicht mehr in diesen Schulprogrammen.

Spalte 5 und 6: Schulen, die sowohl reinen Regelunterricht als auch early-exit TBE anboten und als „Nachbarschafts-Vergleichsschulen“ für one-way DBE-Programme (Spalten 7 und 8) herangezogen wurden.

Spalte 7 und 8: Schulen, die sowohl reinen Regelunterricht als auch one-way DBE-Programme anboten („Nachbarschafts“-Vergleichsgruppe von Spalte 5 und 6).

Spalte 9 und 10: Schulen, die sowohl reinen Regelunterricht als auch early-exit TBE anboten und als „Nachbarschafts-Vergleichsschulen“ für bilinguale two-way-Programme dienten und dabei nach Angaben der Autoren in sozio-ökonomisch besser gestellten Wohnvierteln lagen.

Spalte 11: hier ist für die jeweilige Zeilen angegeben, ob es sich bei den Zahlenangaben um MW = Mittelwert, n = Fallzahl oder SD = Standardabweichung handelt.

Mittelwerte (MW) mit  $n < 30$  sind in Klammern gesetzt; Mittelwerte und Standardabweichung (SD) sind auf eine Dezimalstelle gerundet

Betrachtet man die Daten aus der amtlichen Schulstatistik, erscheint es zunächst offensichtlich, dass (ehemalige) LEP-SchülerInnen, die in ESL-Klassen (Spalte 3) oder vor allem in monolinguale Regelklassen (Spalte 4) gingen, sich in einem immer niedrigeren Leistungsspektrum bewegen, je höher die Jahrgangsstufen werden. Besonders beeindruckend scheinen die steil abfallenden Werte für die Letzteren, die „waivered LEP students“: vom NCE-Mittelwert 49 in der ersten auf 25 in der elften Klasse. Diese amtlichen Daten ebenso wie die ihres eigenen Samples resümieren Thomas und Collier wie folgt: “They are not doing at all well in school. [...] Overall, given the very low achievement district-wide of the waivered LEP students, we would strongly recommend that parents be counseled as to the longterm consequences of their denying their children bilingual/ESL services.” (a.a.O.: 138)

Zunächst sei daran erinnert, dass es sich um Querschnittsdaten handelt. Wenn bei bilingualen Kindern ohne Förderunterricht der Durchschnittswert für Erstklässler wesentlich höher als für Fünftklässler ist, so bedeutet das nicht, dass die Erstklässler in vier Jahren genauso unterdurchschnittlich abschneiden werden, oder umgekehrt die Älteren zu Beginn ihrer Schullaufbahn zu den besseren SchülerInnen zählten. Dies ließe sich nur mit Paneldaten zeigen. Auch könnte es sich theoretisch bei den älteren Jahrgängen (Kohorten) etwa um solche mit einer besonders problematischen Zuwanderungsgeschichte handeln und nicht die Art der Beschulung entscheidend sein. Tatsächlich stellen die Daten also nur Momentaufnahmen für verschiedene Jahrgänge dar – ohne Berücksichtigung einer Reihe einflussreicher Variablen in einem statistischen Analysemodell sind keine Aussagen zu den Ursachen der festgestellten Unterschiede zulässig.

Erste Zweifel an der Aussagekraft der steil abfallenden Werte für diese „waivered LEP students“ in der Schulstatistik sind angebracht, da die Fallzahlen nicht nach Jahrgangsstufen ausgewiesen sind. Zumindest in den von Thomas und Collier selbst analysierten Schülerakten sinkt die Fallzahl ab der sechsten Klasse rapide auf ein für repräsentative Aussagen zu niedriges Niveau ab. Möglicherweise sind also auch bei den amtlichen Statistiken die oberen Jahrgänge dieser Schülergruppe schwach besetzt. Hier setzt auch die irreführende Präsentation der Ergebnisse aus Thomas` und Colliers eigener Stichprobe an. In der schriftlichen Darstellung dieser Ergebnisse (a.a.O.: 137f.) wird der Mittelwert der sechsten Jahrgangsstufe (25. NCE) als Kontrast zum Wert der Erstklässler (48. NCE) genommen, obwohl sie nur die Werte von 25 „waivered“ Sechstklässlern berücksichtigen können und außerdem für ihre bilingualen Vergleichsprogramme gar keine Werte für die sechste Jahrgangsstufe ausweisen. Generell vermeiden die Autoren eine tabellarische Gegenüberstellung dieser Vergleichsgruppen. Dies wurde hier in der Tabelle 5 nachgeholt:

Verglichen werden die NCE-Mittelwerte der Lesetests für die Jahrgangsstufen 1 bis 5 (also die Zeilen) jeweils Spalte 5 mit 6, 7 mit 8 und 9 und 10. Da die Autoren diese durchschnittlichen NCE-Werte für bi- und monolingual unterrichtete SchülerInnen nicht selbst gegenübergestellt haben, liegen auch keine statistischen Signifikanztests dazu vor, ob die vorhandenen Unterschiede statistisch bedeutsam sind. Im Nachhinein lassen sich solche Teststa-

tistiken nicht mehr berechnen.<sup>18</sup> Allerdings sind von diesen 14 Vergleichsmöglichkeiten keine größeren Differenzen als 4,5 NCEs (grob ein Viertel der Standardabweichung) festzustellen – mit Ausnahme eines Vergleichs zwischen Fünftklässlern (5. und 6. Spalte), bei dem die Differenz im Mittelwertvergleich bei 11,1 NCEs liegt. Betrachtet man die große Streuung (durchschnittlich ca. SD  $\approx$  18 NCEs) bei all diesen Mittelwerten, wird deutlich, dass die Leistungsspektren der jeweiligen Vergleichsgruppen sich überwiegend überlappen. Die gelegentlichen marginalen Vorsprünge der bilingual unterrichteten Kinder, die sich in diesen einfachen Querschnittsdaten zeigen, können also keinesfalls als abgesicherte klare Überlegenheit bilingualer Programme interpretiert werden, wie es Thomas und Collier vornehmen. Auf dieser rein deskriptiven Ebene ist für die Leseverständnis- ebenso wie die Mathematik- und Sprachtests die Interpretation, dass meist keine bedeutenden Unterschiede zwischen zweisprachig unterrichteten Kindern und „waivered LEP students“ vorliegen, die zutreffende.

Zudem muss man berücksichtigen, dass es möglicherweise einen Bias in Hinsicht darauf gibt, für welche Kinder ab welcher Jahrgangsstufe überhaupt Ergebnisse aus standardisierten Leistungstests vorliegen: Die amtlichen Schulstatistiken weisen keine Testscores für Erst- und Zweiklässler in early-exit TBE und ESL-Programmen aus. Warum können aber Ergebnisse in der Thomas/Collier-Stichprobe angeführt werden? Rossell und Kuder (2005: 71) verweisen etwa darauf, dass LehrerInnen in bilingualen Programmen ihre SchülerInnen erst dann Leistungstests in englischer Sprache ablegen lassen, wenn deren Englischkenntnisse als gut genug dafür eingeschätzt werden – während in rein englischsprachigen Programmen fast alle Kinder aus Sprachminderheiten diese Tests ablegen. Ohne dies anhand der nötigen statistischen Angaben überprüfen zu können, ist also bei diesen Daten von Thomas und Collier eine eventuell zu positive Darstellung der Leistungen bilingualer Schüler möglich.

**Resümee und Kommentar:** Während die Darstellung der empirischen Vorgehensweise in der 1997er-Studie von Thomas und Collier zu undurchschaubar bleibt, um überhaupt ernst genommen werden zu können (vgl. hierzu auch die ausführliche Kritik von Rossell 1998, sowie Hopf 2005: 240), ist auch in der Veröffentlichung von 2002 keinerlei solide empirische Evidenz für deren Behauptung zu finden, dass es für den Schulverlauf eines LEP-Schülers schädlich sei, nicht an bilingualen Programmen teilzunehmen: Drei von vier Teiluntersuchungen lassen den Vergleich von bi- und monolingual unterrichteten Kindern mit nicht-englischer Muttersprache gar nicht erst zu. Die vierte Teilstudie in Texas hat weder die notwendigen Daten (Prätests), noch wenden die Wissenschaftler angemessene multivariate Analysetechniken an, um ihren postulierten positiven Kausaleffekt bilingualer Programme belegen zu können. Selbst bei einfachen Mittelwertvergleichen neigen die Autoren dazu, die Ergebnisse in ihrem Sinne verzerrend darzustellen. Für einen fairen Vergleich wäre es außerdem angebracht gewesen, auch die Leistungstests von Schülern in englischsprachigen Förderklassen (ESL) systematisch in den Vergleich einzubeziehen. Dies widerspricht offensichtlich dem Interesse der ForscherInnen, den positiven Effekt bilingualer

---

<sup>18</sup> Denn es ist ungeklärt, ob für die Vergleichsgruppen eine Normalverteilung vorliegt, was eine Voraussetzung für einen t-Test wäre, so dass dieser aus den angegebenen Mittelwerten, Standardabweichungen und N nicht errechnet könnte. Auch zur Berechnung eines U-Tests fehlen die Originalwerte der einzelnen Schüler.

Programme „ungeschmälert“ und ohne systematischen Vergleich mit konkurrierenden Förderprogrammen für LEP-SchülerInnen präsentieren zu können. Schließlich suggeriert die Nennung einer Gesamtstichprobe von 210.054 Schülerakten im Executive Summary (a.a.O.: 2) eine Repräsentativität ihrer Aussagen, die angesichts der zahlreichen Mängel und Inkonsistenzen der Studie gänzlich irreführend ist. Insgesamt ist das wissenschaftliche Fundament der Untersuchungen von Thomas und Collier so unsolid, dass sie sich nicht als verlässliche Stütze für ein Plädoyer zugunsten zweisprachiger Unterrichtung eignen.

## 5.3 Studien aus Deutschland und anderen europäischen Ländern

### Deutschland

Zu den ohnehin wenigen bilingualen Unterrichtsmodellen, die an deutschen Grundschulen praktiziert werden<sup>19</sup>, liegen wiederum noch weniger veröffentlichte wissenschaftliche Begleituntersuchungen vor. Diejenigen Studien, die recherchiert werden konnten und die über Beschreibungen von Unterrichtsmodellen hinausgehend auch Sprach- oder Leistungstests erhoben haben, werden im Folgenden kurz vorgestellt werden. Die Durchsicht der Publikationen war dabei auf die spezielle Frage dieses Berichts ausgerichtet, die nach den Effekten von zweisprachigem im Vergleich zu rein deutschem Unterricht auf die Deutschkenntnisse oder Schulleistungen von Kindern nicht-deutscher Herkunftssprache bzw. bilingual aufwachsender Kinder.

**Sandfuchs und Zumhasch (2002)** untersuchen die deutsch-italienische Grundschule in Wolfsburg. Bei diesem Schulmodell handelt es sich um ein two-way-Programm mit Schülern deutscher und italienischer Muttersprache sowie aus deutsch-italienischen Elternhäusern. In halber Klassenstärke wird täglich 30 Minuten Italienischunterricht gegeben, mittlerweile erfolgt eine zweisprachig koordinierte Alphabetisierung (a.a.O.: 104f.). Obwohl sich diese Untersuchung als Evaluation versteht, sind bedauerlicherweise die Voraussetzungen einer aussagekräftigen Evaluierung nur unzureichend erfüllt – zumindest mit Blick auf eine zu messende Effektivität für den deutschen und italienischen Spracherwerb. Zwar ist das Projekt selber längerfristig angelegt – seit 1993 haben vier Jahrgänge den Schulversuch durchlaufen – entgegen der Behauptung der Autoren, dass es sich hier um eine „Längsschnittuntersuchung“ (a.a.O.: 109) handele, fehlten aber für einzelne SchülerInnen im Längsschnitt erhobene Daten. Vielmehr wurden für jeden Jahrgang am Ende der vierten Klasse vier unterschiedliche Typen von Leistungstests durchgeführt. Ohne einen Prätest (und mehrere Retests) fehlt aber die Vergleichsmöglichkeit, von welchem Ausgangsniveau aus die unterschiedlichen Schülergruppen ihre Testergebnisse in der vierten Klasse erreicht haben. Daneben fehlte eine adäquate Vergleichsgruppe nur auf Deutsch unterrichteter SchülerInnen.

---

<sup>19</sup> Für eine Auflistung vgl. Gogolin u.a. (2003: im Anhang 4-8) sowie der Überblick unter <http://lernen.bildung.hessen.de/bilingual/schulverweise/LinkListSchuGrund>; zu pädagogischen Konzepten, Implementationspraktiken und Erfahrungsberichten vgl. auch Fillia (2003), Haase (2003), May (2003) und Kuhs und Steinig (1998).

Der Vergleich von 4.-Klasse-Schulnoten, standardisierten und informellen Leistungstests erfolgte zwischen Kindern mit italienischem, deutsch-italienischem und deutschem Hintergrund – erstere mehrheitlich mit sozio-ökonomisch niedrigem Status, letztere aus der Mittelschicht. Pädagogisches Ziel ist implizit, dass Kinder italienischer Herkunft nicht minderen Erfolg haben sollen. Gleichzeitig waren eben diese zwei Gruppen so unterschiedlich bezüglich Sprache und Schicht, dass es illegitim ist, bei einem Leistungsvergleich am Ende der Grundschulzeit die verbleibenden (erheblichen) Differenzen nur dem speziellen Programm zuzuschreiben.

In allen Tests auf Deutsch sind deutsche Kindern denjenigen italienischer Herkunft im Durchschnitt überlegen. „Mindestens ein Viertel der italienischen Schüler hat – gemessen an den Testergebnissen – einen eindeutigen Förderbedarf im Lesen und Rechtschreiben. Die praktizierte Förderung greift für einen bedeutsamen Teil der italienischen Kinder zu kurz ...“ (Sandfuchs/Zumhasch 2002: 136).

Angesichts der Mittelschichtszugehörigkeit der deutschen Schüler ist es außerdem wenig überraschend, dass diese „allen, den Modellversuch begleitenden Unkenrufen zum Trotz [...] nicht benachteiligt“ wurden (a.a.O.: 135). Eine Kontrollgruppe mit italienischen Kindern an monolingualen Schulen hätte die Möglichkeit eröffnet, fundiert zu beurteilen, ob die italienischen Kinder der Modellschule zwar nicht so gut wie die deutschen KlassenkameradInnen, aber doch besser als italienische Kinder an anderen Schulen abschnitten.

Anhand differenzierter linguistischer Analysen von Schreibproben untersuchte **Roth (2003)** die Textkompetenz von ZweitklässlerInnen einer portugiesisch-deutschen Schulklasse in diesen zwei Sprachen. Wie auch im Wolfsburger Modell fand der Vergleich nur zwischen teilnehmenden SchülerInnen statt, die aus monolingual deutschen ( $n = 7$ ), monolingual portugiesischen ( $n = 5$ ), deutsch-portugiesischen ( $n = 9$ ) und anders zweisprachigen Familien ( $n = 3$ ) stammen. Wegen dieses Untersuchungsschwerpunkts kann die Hauptfrage dieses Berichts nicht beantwortet werden, ob vergleichbare Kinder mit portugiesischer Muttersprache in deutschsprachigen Regel- oder Förderklassen eine bessere Textkompetenz im Deutschen entwickelten. Wegen der fehlenden Vergleichsmöglichkeit ist daher das Ergebnis schwer einzuschätzen, dass es keine statistisch relevante Korrelation zwischen der Gruppenzugehörigkeit der Kinder und den gemessenen Kompetenzen im Deutschen gab: „Die Leistungsstreuung im Deutschen geht nicht so sehr auf das Konto zweisprachiger Kinder, sondern die Leistungen der Gruppen verteilen sich über das Leistungsspektrum der ganzen Lerngruppe ...“ (a.a.O.: 383). Von der kleinen Fallzahl abgesehen, die eine Generalisierung ohnehin verbieten würde, fehlen in dieser Publikation Angaben zum sozio-ökonomischen Hintergrund der Familien und deren Bildungsnähe, um den möglichen Erfolg dieses Modells über den spezifischen Kontext hinaus einschätzen zu können.

Mit sehr ausdifferenzierten Erhebungsinstrumenten hat **Gräfe-Bentzien (2001)** in ihrer Dissertation an einer deutsch-italienischen Europaschule in Berlin die deutsche und italienische Sprachkompetenz von SchülerInnen mit Italienisch als dominanter Muttersprache untersucht. Allerdings war das Forschungsinteresse darauf ausgerichtet, Sprachkenntnisse in L1 und L2 mit der Sprachkompetenz von deutschen Muttersprachlern in Deutschland und italienischen Muttersprachlern in Italien zu vergleichen. Sie kommt zu dem Ergebnis, dass die deutschen Sprachfähigkeiten bei den „Schülern der italienischen Sprachgruppe



gegen Ende des 2. Schuljahres schon fast auf demselben Niveau wie bei der deutschen Sprachgruppe“ (Gräfe-Bentzien 2001: 393) liegen. Allerdings fehlt auch hier der Vergleich mit Schülern italienischer Herkunftssprache, die auf eine deutsche Grundschule ohne Italienischunterricht gehen.

Eine etwas andere Fragestellung untersuchte die wissenschaftliche Begleitung des Berliner Modellprojekts der zweisprachigen, koordinierten Alphabetisierung türkischer Kinder (**Nehr u.a. 1988; Nehr/Karajoli 1995**). An diesem transitorisch angelegten Modell (anfängliche L1-Unterrichtung als Mittel zur Verbesserung des Deutschen) haben türkische Kinder in reinen so genannten „Ausländerklassen“ (n= 36) einerseits und in deutsch-türkisch gemischten Klassen (n= 40) andererseits teilgenommen (im Sinne eines asymmetrischen two-way-Modells, wo die deutschen Kinder (n= 44) aber keinen eigenständigen Türkisch-Unterricht hatten) (1995: 53f).<sup>20</sup> Dies hat den Nachteil, dass der Effekt der zweisprachigen Alphabetisierung nicht von der gemeinsamen Beschulung mit deutschen Muttersprachlern unterschieden werden kann. Da zur Einschulung der L1- und L2-Sprachentwicklungsstand nicht erhoben wurde und offenbar auch sonstige Hintergrundvariablen nicht statistisch kontrolliert wurden, sind darüber hinaus Kausalaussagen auf Basis der Lesetests Ende der ersten und zweiten Klassen nicht zulässig. Dass in bivariaten Ergebnissen die bilingual unterrichteten türkischen Kinder in gemischten Klassen in mehreren Teilstests – auf dem 5%-Niveau signifikant – besser abschnitten als solche in „Ausländerklassen“ und mit deutschen SchülerInnen etwa gleichauf lagen (a.a.O.: 57; 61), ist also kein methodisch gesichertes Argument für zweisprachige Alphabetisierung. Theoretisch plausibel mag die Schlussfolgerung sein, dass eine monolinguale L2-Unterrichtung nur unter L2-Lernern (Ausländerklassen) insofern keine günstige Lernumgebung ist, als ihnen der informelle L2-Erwerb durch Kommunikation mit Muttersprachlern verwehrt ist (vgl. auch a.a.O.: 62).

In einer wissenschaftlichen Untersuchung (**Kupfer-Schreiner 1994**) eines Nürnberger two-way-Modells der gleichzeitigen Alphabetisierung in Deutsch und Spanisch lag das Forschungsinteresse ebenso wie bei den meisten anderen deutschen Untersuchungen auf dem Vergleich zwischen den teilnehmenden SchülerInnen deutscher und nicht-deutscher Muttersprache. Auf Basis einer teils quantifizierten, teils interpretativen Analyse der Textschreibproben von 19 Kindern zu drei Zeitpunkten eines Schuljahres verglich Kupfer-Schreiner die sprachliche Entwicklung dieser Sprachgruppen (a.a.O. 179ff.). Die deutsch-spanisch zweisprachigen Kinder wiesen einen relativ höheren Lernzuwachs auf und holten so von einem niedrigeren Ausgangsniveau gegen über den monolingualen deutschen Muttersprachlern vollständig auf (a.a.O.: 273f.). Eine statistische Kontrolle von weiteren Schülermerkmalen erfolgte nicht.

Zu den wenigen Studien über in Deutschland praktizierte bilinguale Modelle lässt sich für diesen Bericht folgendes Resümee ziehen: Die Forschungsfrage konzentriert sich auf den

---

<sup>20</sup> Zu einem nur kursorisch erwähnten früheren Vergleich von einer zweisprachig unterrichteten türkischen „Ausländerklasse“, einer in Deutsch unterrichteten „Ausländerklasse“ und türkischen Kindern in einer Regelklasse wurde bis Ende der zweiten Klasse kein Programmeffekt auf die Lesekompetenz im Deutschen konstatiert (Nehr/Karajoli 1995: 51).

Vergleich der erworbenen sprachlichen Kompetenzen (im Deutschen und der Partnersprache) der in den bilingualen Klassen unterrichteten Kinder untereinander – also der zur Einschulung monolingual deutschen Muttersprachler, bilingual oder anders monolingualen Kinder. Ohne Zweifel ist die durch detaillierte linguistische Analysen aufgezeigte, teils erfolgreiche Annäherung der Nicht-Muttersprachler an die Sprachkenntnisse bzw. Schulleistungen ihrer monolingual deutschsprachigen KlassenkameradInnen von Belang. Ohne eine Vergleichsgruppe von Kindern nicht-deutscher Muttersprache in rein deutschsprachig unterrichteten Klassen lässt sich damit aber nicht die für diesen Bericht interessierende Frage beantworten, ob bilingualer Unterricht in der Kompetenzvermittlung besser abschneidet als Regelklassen oder – wohl noch relevanter – die Förderung durch systematischen Einsatz von DaZ-Methoden.

Im Rahmen des Forschungsprojekts „BeLesen – Berliner Längsschnittstudie zur Lesekompetenzentwicklung von Grundschulkindern (Schriftspracherwerb von Grundschulkindern nichtdeutscher Herkunft)“ [<http://www.fu-berlin.de/allg.paedagogik/>] unter der Leitung von Hans Merrens arbeiten derzeit Christina Limbird und Andreas Heintze an Analysen, die erstmals einen solchen Vergleich bilingual und monolingual unterrichteter SchülerInnen ermöglichen. 56 an fünf Kreuzberger Schulen zweisprachig unterrichtete Kinder, die wöchentlich ca. sechs Stunden Türkisch-Unterricht erhalten, werden mit 56 nur auf Deutsch unterrichteten Kindern ohne spezielle Förderung verglichen. Die Kinder aus dieser Experimental- und Kontrollgruppe werden auf Basis folgender Merkmale gematcht (paarweise Parallelisierung der Stichprobe): durchschnittlicher sozioökonomischer Status des Schulbezirks (bzw. der geographischen Einheit von Verkehrszellen), Geschlecht, kognitive Fähigkeiten beim Prätest, Sprachehintergrund und Anteil der Kinder mit nicht-deutscher Herkunftssprache in der Klasse. In dieser Panel-Untersuchung werden seit Mitte der ersten Klasse halbjährlich standardisierte Leistungstests erhoben. Zurzeit (Mai 2005) sind die Kinder Mitte der dritten Jahrgangsstufe Klasse. Bei bilingual unterrichteten Kindern werden ihre Rechtschreibung und ihr Leseverständnis im Türkischen und Deutschen getestet, die Kontrollgruppe nur im Deutschen. Zudem werden Mathematiktests durchgeführt.

Eine wissenschaftliche Evaluierung mit systematisch ausgewählten Vergleichsgruppen steht darüber hinaus für den in den Bundesländern unterschiedlich praktizierten muttersprachlichen Ergänzungsunterricht noch vollständig aus (vgl. Hopf 2005: 239).

## **Niederlande**

In den Niederlanden wurde – von einigen zweisprachigen Modellprojekten abgesehen – 35 Jahre lang (bis August 2004) muttersprachlicher Unterricht, *Onderwijs in Eigen Taal en Cultuure* (OETC) bzw. *Onderwijs in Allochtone Levende Talen* (OALT) erteilt. Dieser freiwillige Unterricht umfasste bis zu 2,5 Stunden pro Woche während des vormittäglichen Regelunterrichts und maximal 2,5 Stunden pro Woche zusätzlich am Nachmittag (Driessen 2005: 82) – immerhin bis zu 10 Prozent der gesamten Unterrichtszeit in der 6-jährigen Grundschule.

Zu der Fragestellung dieses Berichts liegen mit Ausnahme Verhoevens (1987) keine deutsch- oder englischsprachigen Publikationen vor. Driessen (2005) stellt die wenigen Studien aus den Niederlanden vor, die in den 1980er Jahren zweisprachigen bzw. muttersprachlichen Unterricht untersucht haben. Diese sechs, mehrheitlich in den 1980ern durchgeführten Evaluationsstudien untersuchten in erster Linie die Auswirkung auf die L1-

Kenntnisse, nur Verhoeven (1987), Driessen und Kollegen (1989) sowie Wagenaar (1993) nehmen auch die Niederländischkenntnisse in den Blick.

Neben einer vertiefenden Analyse des Zweitspracherwerbs bei türkischen Kindern in den Niederlanden vergleicht **Verhoeven (1987)** in seiner (nicht-randomisierten) Untersuchung u.a. den Prozess der Alphabetisierung und der Lese- und Schreibkompetenzen – „literacy processes and literacy outcome“ – zwischen drei Vergleichsgruppen: Kinder, die im reinen Regelunterricht als erstes in L2 alphabetisiert werden (Gruppe 1), und solche in 50:50 transitorischen (Gruppe 2a) bzw. 50:50 simultanen (Gruppe 2b) bilingualen Modellen (Verhoeven 1987: 3). Allerdings entsprach der Regelunterricht insofern nicht der reinen submersion-Variante, als auch diese Kinder im Rahmen des zu der Zeit sehr verbreiteten muttersprachlichen Ergänzungsunterrichts 4 Stunden pro Woche das Fach Türkisch hatten, aber erst ab der zweiten Klasse in dieser Sprache Lesen und Schreiben lernten (a.a.O.: 105, 238; zum sukzessiven bilingualen L2-L1 Modell a.a.O.: 11). Unter statistischer Berücksichtigung der mündlichen L1- und L2-Kenntnisse zu Beginn der Alphabetisierung stellt Verhoeven fest, dass am Ende der zweiten Klasse keine signifikanten Unterschiede im Niederländischen zwischen Experimental- und Kontrollgruppen bestehen (a.a.O.: 249) (die Experimentalgruppe hat auf höherem Niveau angefangen). Allerdings besteht die Experimentalgruppe aus insgesamt nur 25 bilingual unterrichteten Kindern, von denen bis zum letzten Messzeitpunkt sieben aus der Untersuchung ausscheiden, da sie die Klasse wiederholen oder remigrieren. Daher ist eine Aussage zu den Effekten dieser unterschiedlichen Unterrichtsvolumina in der Erstsprache auf den L2-Erwerb nur sehr eingeschränkt möglich.

**Driessen u.a. (1989)** untersuchen mit einem für die Niederlande repräsentativen Sample 12-jährige GrundschülerInnen türkischer und marokkanischer Herkunft. Die erhobenen standardisierter Testscores für Türkisch bzw. Standard-Arabisch, Niederländisch sowie Mathematik (auf Niederländisch) werden jeweils in Bezug auf den bisher besuchten muttersprachlichen Unterricht verglichen, der unterschiedlich häufig und sowohl während als auch nach dem vormittäglichen Regelunterricht erteilt wurde. Zwischen den Leistungen in niederländischer Sprache und dem muttersprachlichen Unterricht besteht nur eine schwache Korrelation, die dann negativ ausfällt, wenn der Zusatzunterricht außerhalb des Regelunterrichts am Nachmittag stattgefunden hatte (Driessen 2005: 94ff.)

**Wagenaar (1993)** präsentiert Driessen (2005: 99ff.) zufolge uneinheitliche Ergebnisse zum Zusammenhang des Erwerbs von Niederländischkenntnissen und dem Besuch einer bilingualen Vorschule: Zwar liegt der Fokus der Studie insgesamt auf der Entwicklung der L1-Kenntnisse. Daneben konstatiert Wagenaar jedoch einen neutralen Effekt der bilingualen Unterrichtung auf L2. Driessen dagegen betont stärker, dass aus der Experimentalgruppe viele marokkanische Kinder – vermutlich aufgrund von schwachen Leistungen in L2 – auf die Sonderschule überwiesen wurden, bevor sie den Posttest ablegen konnten. Dabei handelte es sich vornehmlich um Kinder, die im bilingualen Modell Arabisch gelernt hatten, obwohl sie in den Familien Berber sprachen. Diese Unterrichtung einer vermeintlichen Muttersprache bei zu geringem Gebrauch des Niederländischen scheint hier nicht einer optimalen Förderung der Betroffenen entsprochen zu haben.

Driessen gibt zu bedenken, dass der gängige muttersprachliche Ergänzungsunterricht in den Niederlanden nur eine abgespeckte und von der Unterrichtsqualität zweifelhafte Form der zweisprachigen Erziehung war. Eingedenk der existierenden methodischen Mängel der Studien resümiert er, dass in dieser Implementationsform zweisprachige Erziehung im günstigsten Fall den Verlust der Erstsprache verzögert und keinen Effekt auf die Schulleistungen im Niederländischen hat. Wird den Kindern auch noch zugemutet, neben dem Erwerb des Niederländischen eine Standardsprache aus dem Herkunftsland zu erlernen, statt eine Unterstützung im muttersprachlichen Dialekt zu erfahren (Standard-Arabisch statt Berber), sind eher Nachteile zu erwarten. Ähnliche negative Folgen wären also etwa auch denkbar, wenn in Deutschland Kinder mit kurdischer Muttersprache zusätzlich in Türkisch als vermeintlicher Erstsprache alphabetisiert würden.

## **Schweden**

Obwohl Schweden seit den 1970er Jahren muttersprachlichen Unterricht für SchülerInnen nicht-schwedischer Muttersprache anbietet, ist auch hier die Forschungslage, was die Evaluierung der Effekte betrifft, eher disparat. Nach Axelsson (2005) befasst sich keine quantitativ vorgehende Evaluationsstudie systematisch mit der Frage der Wirkung des muttersprachlichen Unterrichts auf L2 und die Schulleistungen im Allgemeinen. Über Kausalwirkungen des muttersprachlichen Unterrichts kann auch die Studie von **Hill (1995)** keine Aussage machen, die auf einem Workshop der Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI) von Axelsson vorgestellt wurde. In dieser qualitativen Untersuchung über 42 Immigrant\*innen schnitten die SchülerInnen mit einem kontinuierlichen Besuch des herkunftssprachlichen Unterrichts nach eigener Einschätzung und der der Interviewerin in den schulischen Leistungen am besten ab, während die mit nur sporadischem Besuch schlechter Schwedisch sprachen und die weniger erfolgreichen Schulkarrieren hinter sich hatten. Ohne dass eine multivariate statistische Analyse möglich wäre, deutete sich in den Interviews an, dass die elterliche Unterstützung auf das schulische Fortkommen einen erheblichen Einfluss gehabt haben könnte.

## **Exkurs: Studien über Kanada**

Wie relevant sind die Evaluationsstudien zu kanadischen French Immersion-Programmen? In frankophonen Teilen Kanadas gibt es eine Reihe von Schulmodellen für anglophone SchülerInnen, in denen die Fremdsprache Französisch zunächst die dominierende Unterrichtssprache ist und die Kinder nur in dieser L2 alphabetisiert werden, während die Alphabetisierung in Englisch (L1) ab der dritten Klassen (in der Programmvariante „early total immersion“) bzw. in anderen Modellvarianten parallel erfolgt (vgl. Siebert-Ott 1999: 43ff.; als Pionierstudie gilt u.a. Lambert/Tucker 1972). Lassen sich die Erkenntnisse dieser als sehr erfolgreich geltenden Schulmodelle auf die Situation von Kindern aus Zuwandererfamilien oder Sprachminderheiten übertragen?

Rossell plädiert dafür, diese Studien nicht prinzipiell auszuschließen – in ihrem Review mit Baker (1996) hat sie letztlich sechs Studien dieser Art eingeschlossen. Dabei ist ihr Haupt-

argument „brains don't differ“ (Rossell/Kuder 2005: 52), wenn es um Zweitspracherwerb gehe. Doch weder das Argument der grundsätzlichen linguistischen Prozesse noch das der methodischen Fundiertheit der kanadischen Evaluationsstudien kann überzeugen, wenn die Grundkonstellation der Lernenden eine andere ist. Als m.E. wichtigste Argumente halten Slavin und Cheung dem entgegen, dass die Fragestellung eine andere als die ihre (und auch dieses Berichts) sei:

„... the anglophone children are learning a useful second language, not the language for which they will be held accountable in their later schooling. Although most of the studies took place in Montreal, the children lived in English-speaking neighborhoods, and attended schools in an English system. The purpose of bilingual education is to help children succeed in the language in which they will be taught in the later grades, but the French immersion children in Canada are headed to English secondary schools. ... Because French immersion programs were voluntary, children who did not thrive in them could be and were routinely returned to English-only instruction. [...] in many studies, English comparison groups were not learning French at all [...] the focus of the Canadian studies is whether or not French immersion harms the English language development of native English speakers.“ (Slavin/Cheung 2003: 17f.)

Auf die deutsche Situation übertragen wäre die entsprechende Frage, ob etwa der Besuch einer englischsprachigen Diplomatschule der deutschen Sprachentwicklung von deutschsprachigen Kindern schadet und sie außerdem besser Englisch lernen als Gleichaltrige in Regelschulen. Dagegen lässt sich keine vergleichbare Frage z.B. für Kinder türkischer Herkunftssprache in Deutschland formulieren.

Ferner gibt es in den kanadischen Evaluationsstudien nicht dieselbe Art der Kontrollgruppe, wie sie in den bisher vorgestellten US-Studien eine notwendige Bedingung war. „... there is no ‚bilingual‘ group to which immersion could be compared“ (Slavin 2003: 18) Denn die Kontrollgruppen der monolingual englischsprachigen Kanadier gehen in englischsprachige Klassen, bei denen Französisch nur als eine normale Fremdsprache unterrichtet wird und eine Leistungserbringung auf Französisch in keinen weiteren Fächern vorgesehen ist. Eine Leistungserbringung in L2 ist aber dagegen das Hauptziel für Hispanics in den USA oder türkische Kinder in Deutschland. Bei den Studien zu kanadischen Immersion-Programmen fehlt also der Vergleichsmaßstab, wie er bei Studien zu englisch-spanischen Programmen in den USA gegeben ist.

## **5.4 Metaanalysen zur Effektivität zweisprachiger Schulprogramme**

Im Vergleich zur deutschen Forschungslandschaft ist in den USA die wissenschaftliche Tradition der Schuleffektivitätsforschung viel stärker ausgeprägt. Dort stellt sich die Frage, welche der zahlreichen Studien, die auf die eine oder andere Art das Thema dieses Berichts behandeln, methodisch so sauber durchgeführt wurden, dass deren Ergebnisse als wissenschaftlich fundiert angesehen werden können und als Basis für politische Entscheidungen brauchbar sind. Um zu einem zusammenfassenden Urteil zu kommen, wurden in den vergangenen Jahrzehnten mehrere Forschungsüberblicke erarbeitet. Da sich solche Veröffentlichungen vornehmen, zahlreiche Studien zu einer Fragestellung umfassend als state of the art zu resümieren, soll auf die zwei neuen Arbeiten dieser Art von Greene (1998) und Slavin und Cheung (2003 bzw. 2005) nun näher eingegangen werden.

Des Weiteren liegen ältere traditionelle Literaturreviews (Baker/deKanter 1981) und eine frühe Metaanalyse (Willig 1985) vor. Sowohl Greene als auch Slavin nehmen kritisch Bezug auf die Publikation von Rossell und Baker (1996). Rossell und Baker (1996: 23f) kommen zu der Einschätzung, dass Transitional-Bilingual-Education-Programme zwar besser als reine „Submersion“, aber schlechter als Structured Immersion (spezieller Förderunterricht in englischer Sprache mit Englisch als Zweisprache, s. auch Kap. 4) abschneiden. Zu Recht kritisiert Greene einen entscheidenden methodischen Fehler in dem Überblicksartikel von Rossell und Baker: Diese unterscheiden lediglich, wie viele der 72 ausgewählten Studien bzw. Teilergebnisse einen durchschnittlich positiven oder negativen Effekt feststellten („vote counting method“), ohne die Stärke des Effekts in den einzelnen Studien zu berücksichtigen (Greene 1998: 6). Ferner seien die selbst gesetzten methodischen Auswahlkriterien, aufgrund derer Evaluationsstudien als methodisch akzeptabel eingestuft wurden, nicht gründlich überprüft worden. Von den ursprünglichen 72 Studien (aus einem anfänglichen Pool von 300) bei Rossell und Baker stuft Greene nur 11 als akzeptabel ein. Allerdings verfolgten Rossell und Baker insofern eine andere Forschungsfrage, als sie nicht nur an bilingualen Programmen, sondern auch an Vergleichen zwischen rein englischsprachigen Förderprogrammen interessiert waren. Sowohl aufgrund der Kritik an der vote counting method als auch aufgrund der anderen inhaltlichen Fragestellung bei Rossell und Baker (1996) werden im Folgenden nur noch die Ergebnisse und Herangehensweisen von Greene und Slavin berücksichtigt. Allerdings wird die neuste Publikation von Rossell und Kuder (2005) zur kritischen Bewertung der beiden genannten Meta-Analysen mit herangezogen.

Die Besonderheit der Forschungsüberblicke, die Greene und Slavin anbieten, liegt darin, dass es sich um Formen einer Metaanalyse handelt (Slavin/Cheung 2005: 6; zu Metaanalysen allgemein vgl. Beelmann/Bliesener 1994; Cooper/Hedges 1994). Der Anspruch einer Metaanalyse ist es, bei genügend hoher Anzahl an Evaluationsstudien – was hier noch nicht der Fall ist – trotz der Unterschiede in den Programmen, Kontexten, Operationalisierungen und Messinstrumente einen eventuellen Trend in den Effekten systematisch und quantifizierbar aufzudecken. Auf Basis der in den Originalstudien angegebenen Statistiken können so genannte Effektstärken (effect sizes) berechnet werden, die den Vorteil haben, direkt miteinander vergleichbar zu sein, weil sie standardisiert sind (vgl. für einen Überblick Rosenthal 1994: 232-234).

So wird etwa das von Greene (1998: 5) ausgewählte Hedges'  $g$  definiert als die Differenz der Mittelwerte zweier Vergleichsgruppen, dividiert durch die (annähernd unverzerrte Schätzung der) Standardabweichung der Untersuchungspopulation (Rosenthal 1994: 323 sowie Slavin/Cheung 2005: 14). Die Effektgröße ist in Standardabweichungen interpretierbar, d.h. die Messeinheit der Effektgröße sind Standardabweichungen (vom Mittelwert der Vergleichsgruppen). Zwar wird in der Literatur teils vorgeschlagen, Effektgrößen von 0,2 als ‚klein‘, 0,5 als ‚mittel‘ und 0,8 als ‚groß‘ zu bezeichnen, dagegen wird jedoch eingewandt, dass eine solche Interpretation nur im konkreten Kontext inhaltlich Sinn macht (vgl. Beelmann/Bliesener 1994: 222).

## Greene (1998)

Greene untersucht auf Basis systematisch ausgesuchter Evaluationsstudien bilingualer Unterrichtsmodelle deren durchschnittliche Effekte auf die Ergebnisse standardisierter Leistungstests. Für Lesetests auf Englisch, englische Sprachtests bzw. „for all tests in English“ (1998: 5)<sup>21</sup> und, wo vorhanden, für Mathematik-Tests auf Englisch und Spanisch-Tests (als

<sup>21</sup> Zum Teil handelt es sich hierbei um eine nochmalige Aufführung des Lesetests, zum Teil um einen Durchschnittswert aller englischsprachigen Tests inklusive Mathematiktests, was von Rossell und Kuder (2005: 61) als gänzlich unüblich kritisiert wird.

die üblicherweise unterrichtete L1) errechnet Greene Effektgrößen. Greene gibt in seinen Tabellen pro Studie neben der Effektgröße den zugehörigen z-Wert an, aus dem sich (bei  $z > 1,96$ ) ein 5%-Signifikanzniveau ableiten lässt, also die Irrtumswahrscheinlichkeit, mit der sich die Nullhypothese verwerfen lässt, dass der Unterschied zwischen Experimental- und Kontrollgruppe nur zufällig sei. Eine Effektgröße von knapp einem Fünftel entspricht etwa einem zusätzlichen Leistungszuwachs von drei Monaten über eine Zeit von zwei Schuljahren, so Greene (1998: 2).

Tab. 6: Greene 1998: Einzelne Effektgrößen für bilinguale Programme in elf Originalstudien

Studie	Engl. Sprachtest oder Durchschnitt aller engl. Tests		Engl. Lesetest		biling. Programm	Kontrollgruppe
	Effektgröße	z-Wert	Effektgröße	z-Wert	n	n
Bacon 1982	0,79	2,39**	0,68	2,07**	18	18
Covey 1973	0,34	2,94**	0,74	4,87**	86	89
Danoff 1977	-0,03	-0,39	-0,12	-1,50	955	523
Huzar 1973	0,18	0,83	0,18	0,83	43	43
Kaufman 1968	0,20	0,72	0,20	0,72	43	31
Plante 1976	0,52	1,34	0,52	1,34	16	12
Powers 1978	0,001	0,1	-0,33	-1,53	44	43
Ramirez 1991	0,01	0,08	0,12	0,73	88	160
Rossell 1990	-0,01	0,03	-0,05	-0,20	174	173
Rothfarb 1990	0,02	0,24	-	-	70	49
Skoczylas 1972	-0,05	-0,18	0,13	0,46	25	25

Quelle: vgl. Greene 1998: Tab. 3

Anmerkung: \*\* signifikant auf dem 5%-Niveau

Zu diesen angeführten Studien ist zunächst grundsätzlich zu bemerken, dass nach eigener Durchsicht nicht ersichtlich ist, warum Greene diese Studien als methodisch akzeptabel bezeichnet. Wie oben (S. 26-27) bereits angeführt, weisen fünf der Studien erhebliche methodische Mängel und Unklarheiten auf (Covey 1973, Danoff 1977, Powers 1978, Rossell 1990, Skoczylas 1972), und drei untersuchen so wenig vergleichbare Populationen (Bacon 1982), Jahrgangsstufen (Kaufman 1986) oder nur mündliche Sprachkompetenzen (Rothfarb u.a. 1987), dass ihre Vermengung mit den anderen Studien inhaltlich nicht gerechtfertigt erscheint. Bei den methodisch besseren Studien von Ramirez u.a. (1991), Huzar (1973) und Plante (1976) stimmt zwar die „grobe Richtung“ der Effektgrößen, d.h. die Differenzen zwischen Kontroll- und Experimentalgruppe sind nicht statistisch signifikant. Aber für die Ramirez-Studie wird nicht die entscheidende Teilanalyse verwendet (s. oben, S. 32-33), bei Huzar (s. oben, S. 35) und Plante wird – ohne dass Greene selber dies offen legt – eine durchschnittliche Effektgröße für (im Programmeffekt) sehr unterschiedliche Kohorten angegeben, was insofern nicht gerechtfertigt ist, als bei diesen randomisierten Studien nicht untersucht wurde, in welchen Merkmalen sich die jeweiligen Jahrgänge in Experimental- und Kontrollgruppe unterscheiden könnten.<sup>22</sup>

Auf einer methodisch so unsicheren Basis erscheint es wenig gerechtfertigt, wenn nicht irreführend, dass Greene eine durchschnittliche Effektgröße seiner elf Studien pro Testart

<sup>22</sup> Deshalb werden auch im Review von Rossell und Baker (1996) solche Studien mit evtl. gegenläufigen Ergebnissen für unterschiedlichen Kohorten mehrfach aufgeführt. „We believe that a study is only redundant if it is of the exact same students in the same year with the exact same tests.“ (Rossell/Kuder 2005: 58)

(für Sprachtests etwa 0,18, für Lesetests 0,21) und dazugehörige durchschnittliche (hier signifikanten) z-Werte berechnet (a.a.O.: Tab.1). Unabhängig von den Mängeln in den Originalstudien ist bei nur 11 Studien, also einer „Stichprobe“  $N = 11$ , das Errechnen eines arithmetischen Mittels irreführend, da sich hier das Problem der statistischen Ausreißer gravierend stellt: So haben bei Greene von elf Studien bei Lesetests nur zwei (Covey 1973; Bacon u.a. 1982) einen starken Effekt mit signifikantem z-Wert. Nur aufgrund dieser zwei Studien, kommt es überhaupt zu dem moderat hohen und signifikanten Durchschnittswert von 21% einer Standardabweichung. Gewichtet man diese elf Studien mit ihren Stichproben (um denen mit kleinen Samples entsprechend weniger Gewicht zu geben und umgekehrt), so würde selbst auf der zweifelhaften Basis seiner Effektgrößen kein signifikanter Unterschied zwischen den Programmtypen herauskommen (Rossell/Kuder 2005: 61 und Tab. 2 und 3).

Letztlich bieten also die zahlreichen Ungereimtheiten und Mängel seiner Metaanalyse (für eine ausführliche Kritik von Greene vgl. Rossell/Kuder 2005: 59ff.) keineswegs die wissenschaftlich fundierte Grundlage für folgenden Schluss: „Despite the relatively small number of studies, the strength and consistency of these results [...] increases confidence in the conclusions that bilingual programs are effective at increasing standardized test scores measured in English.“ (Greene 1998: 6) Bei einer methodenkritischen Durchsicht seiner Analyse sind jedoch mehr Hinweise auf einen neutralen statt einen positiven Effekt bilingualer Unterrichtung auf L2 auszumachen, wobei sich Greenes Metanalyse insgesamt als wissenschaftlich nicht einwandfrei herausgestellt hat.

### **Slavin und Cheung (2005)**

Slavin und Cheung geben ähnlich wie Greene einen Überblick über Programmeffekte, die methodisch akzeptable Evaluationsstudien beim Vergleich der Schulleistungen von bilingual im Vergleich zu monolingual unterrichteten Kindern festgestellt haben. Sie beschränken sich auf Testergebnisse, die für alle dieser Art von Studien vorliegen, d.h. auf die der Lesetests. Dabei unterscheiden sie die Studien nach methodischen und inhaltlichen Kriterien:

1. a) Studien zu bilingualen Programmen, die gleichzeitig alphabetisieren; geordnet nach

- Längsschnittstudien mit Randomisierung
- nicht-randomisierte Längsschnittstudien, die wichtige Drittvariablen kontrollieren bzw. die Experimental- und Kontrollgruppe daraufhin gematcht haben („matched control“);

1. b) Studien, die im Gegensatz zu 1.) und 2.) die Leistungszuwächse nach nur einem Jahr messen;



Tab. 7: Slavin/Cheung 2005: Lesetest-Vergleich zwischen bilingualen und monolingualen Schulprogrammen: deskriptive Informationen und Effektgrößen ausgewählter Evaluationsstudien

Study	Intervention description	Design	Duration	N	Grade	Sample Characteristics	Evidence of Initial Equality	Posttest	Effect Size	Mean Weighted Effect Size
<b>Studies of paired bilingual education</b>										
Plante (1976)	Paired bilingual	Random assignment	2 yrs	55 <sup>1</sup>	1-2, 2-3	Spanish-dominant Puerto Rican students in New Haven, CT	Well matched on Spanish oral vocabulary but C > E in English pretest	English Inter-American Series 2nd grade 3rd grade	+0.78* +0.26	+0.50
Huzar (1973)	Paired bilingual	Random assignment	2 & 3 yrs	160	1-2, 1-3	Disadvantaged Puerto Rican students in Perth Amboy, NJ	Well matched on IQ, SES, and initial achievement	English Inter-American Series 2nd grade 3rd grade	+0.01 +0.31	+0.16
Campeau et al (1975) Corpus Christi	Paired bilingual	Matched control	2 yrs	171 <sup>3</sup>	K-1	Spanish dominant students in Corpus Christi, Texas	Matched on English and Spanish pretests	English Inter-American Series	+0.45*	+0.45*
Ramirez et al (1991)	Early Exit TBE <sup>2</sup> vs Immersion	Matched control	2 yrs	153 students in 4 schools	K-1	Spanish dominant LEP students	Very well matched on SES and home backgrounds.	English CTBS	+0.53*	+0.53*
Campeau et al (1975) Houston	Paired bilingual	Matched control	3 yrs	206	K-2	Spanish dominant students in Houston, TX	Matched on language, SES, and academic achievement	English Inter-American Series	+0.49*	+0.49*
Maldonado (1977)	Paired bilingual	Matched control	5 yrs	126	1-5	Spanish dominant students in six elementary school in Corpus Christi, TX	Matched on SES and number of years in schools	English (SRAAS) 2nd 3rd 4th 5th	0.00 <sup>a</sup> 0.00 <sup>a</sup> 0.00 <sup>a</sup> 0.00 <sup>a</sup>	0.00 <sup>a</sup>
Alvarez (1975)	Paired bilingual	Matched control	2 yrs	147	2	Spanish dominant children in two schools in Austin Texas	Matched on SES and initial language proficiency	California Achievement Tests English reading vocab English reading comp	+0.12 -0.23	-0.06

Cohen (1975)	Paired bilingual	Matched control	2-3 yrs	90	K-1, 1-2, 1-3	Spanish dominant students in Redwood city, CA	Matched on SES and initial language proficiency	English Inter-American Series Cohort 1 Cohort 2	0.00 <sup>a</sup> 0.00 <sup>a</sup>	0.00 <sup>a</sup>
Campeau et al (1975) Kingsville	Paired bilingual	Matched control	1 yr	89	K	Spanish dominant students in Kingsville, TX	Matched on SES and ethnic mix	English Inter-American Series	+0.42*	+0.42*
Campeau et al (1975) Santa Fe	Paired bilingual	Matched control	1 yr	77	1	Hispanic students in Santa Fe, New Mexico	Matched on Pretests, but E > C	English MAT	+0.03	+0.03
<b>Studies of one-year transitional bilingual education</b>										
Maldonado (1994)	Bilingual 1-year-transition in special education	Random assignment	3 yrs	20	2-4, 3-5	Spanish dominant special education students in Houston TX	Well matched on disability, language proficiency, & family background	English CTBS	+1.66*	+1.66*
Campeau et al (1975) Alice	Bilingual 1-year-transition	Matched control	2 yrs	125	K-1	Spanish dominant students in Alice, Texas	Similar on English pretests but E > C on Spanish pretest	English Inter-American Series	+1.06*	+1.06*
<b>Study of bilingual education (unspecified)</b>										
Saldate et al (1985)	Unspecified bilingual instruction	Matched control	3 yrs	38 <sup>3</sup>	1-3	Spanish dominant students in Douglas, AZ	Well matched on pretests	English tests MAT (2nd grade) WRAT (3rd grade)	-0.28 +0.89*	+0.89*

Quelle: vgl. Slavin/Cheung 2005: Tab. 1. Im Unterschied zur bisher zugänglichen Publikation von 2003 wurden die Ordnung der Studien und die Kurzbeschreibung der bilingualen Programme geändert sowie z.T. Effektgrößen neu berechnet; eine Studie von Carlisle und Beeman (2000) ist dort wegen inadäquater Vergleichsgruppen inzwischen ausgeschlossen worden.

Anmerkungen: 1) richtig ist N = 53 für beide Kohorten zusammen

2) In der aktuellsten Version haben Slavin und Cheung die von Ramirez u.a. untersuchte Unterrichtsform als „paired bilingual“ rekategorisiert.

3) möglicher Bias bei Verkleinerung der Stichprobe (attrition) (Slavin/Cheung 2003: 14f.)

a) Die Effektgröße wurde auf Null geschätzt, da die zur Berechnung nötigen Parameter nicht angegeben waren, aber in den Studien von nicht-signifikanten Ergebnissen gesprochen wurde.

\* Nach Angaben von Slavin/Cheung in ihren Studienbeschreibungen waren diese Effektgrößen (vermutlich auf dem 5%-Niveau) signifikant unterschiedlich von Null, so dass von ‚keinem Effekt‘ der bilingualen Unterrichtung gesprochen werden kann; z.T. musste der Ausdruck „substantially higher/better“ als äquivalent zu statistisch signifikant interpretiert werden.

2. Studien, die transitorische Modelle auswerten, in denen bilingualer Kinder nach einem Jahr in den Regelunterricht überwiesen werden;
3. Studien, in denen das bilingualer Programm nicht ausführlich genug beschrieben wird, um es den bisher genannten Modellen zuordnen zu können;
4. Heritage-Language-Programme, in denen die unterrichtete Erstsprache nicht Spanisch ist (sondern Französisch oder eine indianische Sprache, die aber nicht unbedingt L1 der Schüler ist);<sup>23</sup>
5. Studien, die Leistungen in der Sekundarstufe statt wie bei 1.), 2.) und 3.) in der Primarstufe messen.

Die Mehrzahl der evaluierten Programme – „paired bilingual“ Strategien – beinhalten eine gleichzeitige, koordinierte Alphabetisierung: „...students were taught to read in English and in Spanish at different times of the day, beginning in kindergarten or first grade and continuing through the end of the study. Pairing may not have begun on the first day of the school year, but if children were being taught to read in both Spanish and English during their first year of reading instruction, the program was considered a paired model“ (Slavin/Cheung 2005: 17) Gleichzeitige Alphabetisierung ist für zweisprachige Modellklassen in Deutschland typisch, während für die USA selbst die große Mehrheit der bilingualen Programme transitorische early-exit Modelle sind.

Tabelle 7 ist auf die englisch-spanischen Programme in der Primarstufe beschränkt.

Nachdem in Kapitel 5.2 drei der berücksichtigten Studien bereits ausführlich vorgestellt wurden, sollen hier noch folgende Punkte zur abgedruckten tabellarischen Darstellung ergänzt werden: Der größte Mangel dieser Übersicht ist, dass die Fallzahlen pro Studie summiert werden, auch wenn dort mehrere Kohorten (also Jahrgänge mit z.T. unterschiedlich langem Besuch eines bilingualen Programms) mit entsprechend geringeren Fallzahlen untersucht werden, die z.T. wie bei Alvarez (1975) nicht vergleichbar sind bzw. nicht auf systematische Unterschiede hin untersucht wurden (vgl. Plante 1976, Huzar 1973). Ferner ist nicht ersichtlich, warum die beiden Autoren nur mehr oder minder präzise Angaben im Text machen, ob der Effekt des Treatments statistisch signifikant unterschiedlich von Null ist, statt wie Greene (1998) z-Werte oder noch anschaulicher bei Rossell und Kuder (2005: Anhang 6) Konfidenzintervalle zu errechnen.

Von den 13 Evaluationsstudien zu bilingualen Grundschulprogrammen stimmen bei acht die statistische Signifikanz, die die Verfasserin aus Slavin und Cheung (2005) deduziert hat, mit den neu errechneten Angaben von Rossell und Kuder (2005) überein. Saldade u.a. (1985) wird aufgrund des hohen Stichprobenschwunds fehlende Signifikanz attestiert. Zudem stufen Rossell und Kuder von den fünf Teilstudien von Campeau u.a. (1975) nur diejenige über Corpus Christi als wissenschaftlich akzeptabel ein bzw. konnten nicht die notwendigen statistischen Parameter auffindig machen, so dass nach ihren Angaben auch keine Effektgröße berechnet werden sollte. Deshalb sollten die Untersuchungen von Campeau u.a. (1975) weniger Gewicht zugemessen werden, als dies bei Slavin und Cheung der Fall ist. – Dabei merken Slavin und Cheung ohnehin an, dass die American Institutes of Research hier nur, „chosen on their reputation for excellence“ (2003: 14), vorbildhafte bilingualer Programme evaluieren ließen, so dass ein positiver Effekt kaum überrascht. – Ohne diese Studien überwiegt ein neutraler Effekt bilingualen Unterrichts auf die Testergebnisse.

---

<sup>23</sup> nur in Slavin/Cheung (2003)

Im Vergleich zu Greenes Metaanalyse bleiben Slavin und Greene in ihrer Darstellungsweise etwas transparenter, weil sie insbesondere durch die Kurzdarstellungen der Studien methodische Einschränkungen und programmspezifische Besonderheiten aufzeigen. Dadurch wird ersichtlich, dass auch diejenigen Studien, die die oben (S. 21-22) erläuterten Mindeststandards erfüllen, immer noch methodische Fallstricke aufweisen können. Während Slavin und Cheung in der älteren Version (2003) ihrer Publikation darauf verzichteten, eine zusammenfassende durchschnittliche Effektgröße anzugeben, geben sie in der jüngsten Veröffentlichung für die 13 Studien, die bilinguale Grundschulprogramme für Kinder mit spanischer Muttersprache evaluieren, den Median aller nach Samplegröße gewichteten, durchschnittlichen Effektgrößen an: „The mean sample size-weighted effect size for the 13 studies of elementary reading for Spanish-dominant students was +0.33. Using procedures described by Lipsey & Wilson (2001), this effect was found to be significantly different from zero,  $Q=29.6$ ,  $p<.05$ ,  $df=12$ .“ (Slavin/Cheung 2005: 31) Auch wenn der Median der angemesseneren Parameter ist (als das arithmetische Mittel bei Greene), kann man auch hier darüber streiten, ob die Präsentation einer solchen „alles zusammenfassenden“ Zahl angesichts der wenigen, wirklich verlässlich evaluierenden Studien tatsächlich angebracht ist.

Mit der wichtigen Einschränkung, dass es insgesamt viel zu wenige Evaluationsstudien mit hoher methodischer Qualität gibt, resümieren Slavin und Cheung: „... research on language of instruction does yield some important lessons at least worthy of further study. Across 17 qualifying studies of all types of programs, 12 found effects favoring bilingual education and 5 found no differences. None of the studies found results favoring English immersion.“ (2005: 31) Angesichts der Art der bilingualen Programme, die überhaupt (mehr oder weniger) methodisch adäquat evaluiert und z.T. als erfolgreich bewertet wurden, favorisieren Slavin und Cheung Modelle der gleichzeitigen Alphabetisierung: „Teaching reading in two languages, with appropriate adaptations of the English program for the needs of English language learners, may represent a satisfactory resolution to the acrimonious debates about bilingual education.“ (2005: 33)

Aus Sicht der Verfasserin sollte aufgrund der bestehenden methodischen Mängel sowohl in den Originalstudien als auch der Metaanalyse selbst eine Schlussfolgerung noch vorsichtiger ausfallen. Die „Best-Evidence Synthesis“ von Slavin und Cheung lässt die Aussage zulässig erscheinen, dass keine Anzeichen vorliegen, dass sich bilingualer Unterricht negativ auf die Leseleistungen in der Zweitsprache auswirkt. Die Kritiker solcher Programme haben also nicht Recht, wenn sie eine schädliche Wirkung befürchten. Umgekehrt zieht sich ein Vorsprung der zweisprachig unterrichteten SchülerInnen nicht konsequent durch, ein neutraler Effekt ist nicht minder wahrscheinlich (vgl. die Reanalyse bei Rossell/Kuder 2005: Tabelle 4). Vor allem aber liegt im Gegensatz zu Cummins Schwellenhypothese bisher keine Evidenz vor, dass es für den Zweitspracherwerb und die Schulleistungen unabdingbar oder günstiger ist, wenn eine Alphabetisierung zunächst in L1 und erst zu einem späteren Zeitpunkt in L2 erfolgen muss. Dass in beiden Metaanalysen mehrheitlich Studien zur parallelen L1-L2-Alphabetisierung berücksichtigt werden, ist für den US-Kontext eher

unbefriedigend, da dort ja transitorische Modelle die viel üblicheren gewesen sind.<sup>24</sup> Für das Interesse in Deutschland an two-way Modellen mit gleichzeitigen Alphabetisierungen sind sie – unbeabsichtigterweise – relevanter, auch wenn die eindeutigen Präferenzen von Greene sowie Slavin und Cheung für bilinguale Modelle mit ihren eigenen Daten nicht ausreichend belegt sind.

Insgesamt bleibt die wohl unstrittigste Schlussfolgerung aus den beiden Metaanalysen, dass – in den USA und besonders in Deutschland selbst – noch ein großer Bedarf besteht, Modelle der bilingualen Unterrichtung methodisch angemessen zu evaluieren. Im Folgenden soll ergänzend zu Kapitel 5.1 nun skizziert werden, wie eine Evaluationsstudie bilingualer Grundschulprogramme in Deutschland aussehen könnte und welche alternativen Forschungsdesigns möglich und sinnvoll wären.

So sehr die Auswirkung zweisprachigen Unterrichts auf L2 wissenschaftlich umstritten bleibt, so herrscht in Bezug auf dessen Einfluss auf die Entwicklung der Erstsprache mehr Einigkeit. Selbst zwischen US-amerikanischen Befürwortern und Kritikern bilingualer Schulprogramme ist es Konsens, dass solche Modelle die (insbesondere schriftsprachlichen) L1-Kompetenzen der dort beschulten Kinder im Vergleich zu denen im Regelunterricht fördern (Rossell/Kuder 2005: 48; Slavin/Cheung 2005: 23, 27, 30; Greene 1998: Tab. 1 und 3). Allerdings deuten Forschungsergebnisse aus den Niederlanden (Driessen 2005) an, dass auch hier die spezifischen Lernbedingungen den Ausschlag geben können und ein größeres Unterrichtsvolumen in L1 nicht notwendigerweise mit besseren L1-Kompetenzen korreliert (Driessen 2005: 93 zu Verhoevens 1987). Die positiven Effekte der L1-Unterrichtung auf die Erstsprache wirken nicht immer nachhaltig, d.h. die Kompetenzen in der Erstsprache stagnieren, wenn die Beschulung in L1 ausgesetzt wird (Wagenaar 1993). Zumindest der gängige muttersprachliche Ergänzungsunterricht kann den Verlust der Muttersprache (im Zuge sprachlicher Assimilationsprozesse) Driessen zufolge höchstens aufhalten. „Starke“ bilinguale Modelle, die bis in die oberen Jahrgangsstufen an der Ausbildung einer academic language auch in der Erstsprache arbeiten, sind bei einer entsprechenden Zielsetzung sicherlich vielversprechender.

---

<sup>24</sup> vgl. Rossell/Kuder (2005: 69f.), bei denen auf Basis ihrer vote-counting-Methode die „klassischen“ early-exit transitorischen Modelle der monolingualen Fördermethode Structured Immersion eindeutig unterlegen sind (a.a.O.: Tabelle 1b), auch wenn gelte: „Spanish bilingual education in the U.S. is not a disaster“ (a.a.O.: 74).

## 6. Überlegungen zu einer Evaluation von Modellversuchen zweisprachiger Unterrichtung in Deutschland

Eine grundsätzliche Erläuterung der Merkmale einer aussagekräftigen Evaluierung bilingualer Unterrichtsmodelle und typischer methodischer Probleme in diesem Feld empirischer Schulforschung erfolgte bereits in Kapitel 5.1. Im Folgenden sollen Besonderheiten des deutschen Kontexts angesprochen werden.

Im methodischen Vorgehen können sich ForscherInnen in Deutschland nur zum Teil an die langjährige amerikanische Forschungserfahrung anlehnen. Vor allem in Bezug auf die verfügbaren Messinstrumente ist die Differenz offensichtlich. Regelmäßig erhobene, standardisierte Leistungstests, auf denen Evaluatoren im US-Kontext aufbauen, sind für die Bundesrepublik noch nicht verfügbar. Die Durchführung von prospektiven, quasi-experimentell angelegten Evaluationsstudien in Deutschland ist deshalb insbesondere auf die Neu- und Weiterentwicklung von Schulleistungstests und Sprachstandserhebungen angewiesen (vgl. hier die Expertisen zu Verfahren der Sprachstandsfeststellung von Ehlich u.a. 2004, sowie Fried 2004).

Für Deutschland wird die Entwicklung solcher Instrumente für eine Sprachstandserhebung insgesamt noch als unzureichend bewertet (Ehlich u.a. 2004: 5). Von den systematisch überprüften Verfahren zur Sprachstandsfeststellung berücksichtigen „HAVAS-5“ die Herkunftssprache gleichberechtigt, „SDF“ und „sismik“ nur eingeschränkt (a.a.O.: 18, sowie eine ausführlichere Diskussion dieser Tests bei Fried 2004: 33ff., 42ff., 83ff. sowie weiterer Sprachentwicklungstests 49ff.). Wollte man die Wirkung bilingualer Programme über die ersten Grundschuljahre hinweg untersuchen, bedürfte es aber eines Posttests, der auch für spätere Jahrgangsstufen entwickelt wurde. Für diese Jahrgangsstufen gibt es insgesamt wenige (standardisierte Lese- und Rechtschreib-)Tests, wie etwa SDF (Ehlich 2004: 30-32).

Soll die Messung fachlicher Kompetenzen im Mittelpunkt einer Evaluation stehen, so bieten sich als Posttest etablierte Leistungstests an, wie sie etwa in der IGLU-Studie (Bos u.a. 2003: 14f.) oder PISA (zu Tests und dem Konzept von "literacy" in PISA vgl. OECD 2002) zum Einsatz kamen. Kenntnisse der Testsprache sind dabei eine notwendige Voraussetzung, die Kompetenzmessung geht aber darüber hinaus.

Bei der Auswahl der Stichprobe bleibt, wie in Kapitel 5.1 bereits erläutert, zu beachten, dass eine Randomisierung insbesondere aufgrund der Zustimmungspflicht der Eltern kaum zu realisieren ist. Bei quasi-experimentellen Designs aber lässt sich das Problem des Selektionsbias – dass bestimmte Elterntypen für ihr Kind eine bilinguale Unterrichtung wählen – nicht grundsätzlich vermeiden. Zwar wäre es günstiger, wenn die Schülergruppe im bilingualen Programm zumindest hinsichtlich ihrer sozialen Herkunft nicht zu einseitig zusammengesetzt wäre. Die Steuerungsmöglichkeiten sind hier allerdings begrenzt.

Deshalb sind zunächst auf der Ebene der individuellen Merkmale der SchülerInnen bzw. ihres Elternhauses neben einer Form von Prätest die wichtigsten potentiellen Einflussfaktoren auf Zweitspracherwerb und Schulleistungen zu erheben:

- die eigene bzw. die elterliche Migrationsbiografie (z.B. Einreisalter bzw. Aufenthaltsdauer, Migrationsmotiv – Flucht und Asyl versus Arbeitsmigration bzw. Familienzusammenführung, Herkunftsland),
- der sozio-ökonomische Status der Eltern (Einkommen, Erwerbsstatus, berufliche Stellung etc.),
- das kulturelle Kapital der Eltern (z.B. Bildungsniveau, bewährte Indikatoren wie die Anzahl der Bücher),
- das innerfamiliäre soziale Kapital wie z.B. die elterliche Unterstützung der Kinder bei schulischen Aufgaben, die Intensität der Kommunikation innerhalb der Familie
- der Sprachgebrauch von Deutsch versus L1 in unterschiedlichen Kontexten (Familie, Freunde, Medienkonsum).
- Setzt sich die Experimentalgruppe aus ähnlich unterrichteten bilingualen Klassen aber mit unterschiedlichen Partnersprachen zusammen, muss zudem natürlich die Muttersprache erhoben werden, die unterschiedliche linguistische Distanz zum Deutschen aufweisen mag.

Durch die Erhebung dieser Merkmale werden – bei einem genügend großen Sample von teilnehmenden SchülerInnen – differenzierte Aussagen möglich, für welche Teilgruppen der Kinder nicht-deutscher Muttersprache sich bilinguale oder andere Förderprogramme als wie effektiv erweisen. Zudem kann so mit multivariaten Analyseverfahren (Regressionsanalysen) wie etwa bei Ramirez u.a. (1991) der relative Einfluss des zweisprachigen Unterrichts auf den Kompetenzerwerb mit dem dieser anderen Faktoren direkt verglichen werden.

Neben den Eigenschaften der teilnehmenden SchülerInnen muss die Wirkung der Kontexte berücksichtigt werden. Um die örtliche Umgebung und den schulischen Kontext zwischen Experimental- und Kontrollgruppe annähernd vergleichbar zu halten (damit insbesondere nicht Wirkungen der allgemeine Eigenschaften der Schule mit denen des Modellprogramms konfundieren), wäre es wie bei Ramirez u.a. (1991) empfehlenswert, bilinguale mit monolingualen Zügen (Parallelklassen einer Jahrgangsstufe) an jeweils einer Schule zu vergleichen. Wenn das Projekt eines bilingualen Unterrichts noch nicht begonnen hat, bietet sich als alternatives Forschungsdesign auch das Kohorten-Design an (vgl. Cook/Campbell 1979: 126-133). Die erste Kohorte (ab der ersten Klasse) nimmt dabei an dem Programm noch nicht teil, legt aber auch schon Prä- und später Posttests ab und kann so als Kontrollgruppe für die nachrückenden, bilingual unterrichteten Jahrgänge (ebenfalls ab der ersten Klasse) dienen. Hier könnte die Vergleichbarkeit der Experimental- und Kontrollgruppen noch dadurch erhöht sein, dass entsprechend verschieden beschulte Geschwister unterschiedlichen Alters einbezogen sind. Als nächstbeste Alternative müssten SchülerInnen der Kontrollgruppe in derselben oder aufgrund von Sozialindikatoren vergleichbaren Nachbarschaft zur Schule gehen bzw. durch Matching eben solche monolingual unterrichteten Kinder ausgewählt werden (vgl. hierzu das laufende Projekt BeLesen, S. 49).

Um die Verallgemeinerbarkeit und die statistische Aussagekraft einer solchen Untersuchung zu gewährleisten, wäre es wünschenswert, die Untersuchung zumindest auf mehrere bilinguale bzw. Kontrollklassen auszudehnen. Dann sollte wiederum das bilinguale Programm selbst von seiner Konzeption und Umsetzung her möglichst vergleichbar sein (z.B.

die bilingualen Grundschulen in Hamburg<sup>25</sup>). Die Varianz von Kontextmerkmalen (z.B. die Ausstattung der Schulen, Qualifikation der LehrerInnen, sozio-ökonomischer Status und ethnische Zusammensetzung der Klasse und/oder Nachbarschaft) könnte bei einer genügend hohen Anzahl von bi- und monolingualen Klassen so in einem Mehrebenenmodell unabhängig vom Programmeffekt statistisch berücksichtigt werden. In Deutschland erscheint so ein Untersuchungsdesign jedoch kaum realistisch, weil es insgesamt wenige bilinguale Grundschulen bzw. Schulen mit bilingualen Zügen bisher gibt, bei denen eine typische Migrantensprache die Partnersprache ist.<sup>26</sup>

Aus einer problemorientierten Fragestellung heraus kann es Sinn machen, zumindest hinsichtlich des sozio-ökonomischen Status von vornherein eingrenzend danach zu fragen, wie solche Unterrichtsmodelle in gewissen typischen Kontexten, nämlich in urbanen, sozial schwächeren und ethnisch gemischten Vierteln wirken. Denn gerade hier wohnen viele derjenigen Kinder nicht-deutscher Muttersprache, die potentiellen schulischen Förderbedarf haben. Diesen Fokus haben auch die Mehrheit der US-amerikanischen Modelle und Evaluationen, bei denen viele Kinder mit spanischer Muttersprache oft aus armen Familien stammen und die Wohnviertel insgesamt von einem niedrigen sozio-ökonomischen Status und ethnischer Segregation geprägt sind.

Schließlich bleibt zu diskutieren, mit welcher Unterrichtsform bilinguale Modelle verglichen werden sollten. Angesichts der großen Leerstellen in der empirischen Schulforschung in Deutschland sollte eine neue Evaluationsstudie sinnvoller- und idealerweise nicht nur Kontrollgruppen in monolingualen Regelklassen im Sinne von „submersion“ untersuchen, sondern auch deutschsprachige Förderprogramme testen, z.B. den systematischen Einsatz von Deutsch als Zweitsprache, sprachsensiblen Fachunterricht oder „content based language learning“, d.h. sprachliches Lernen an fachlichen Inhalten (vgl. Reich/Roth 2002: 38). Bei so einer vergleichenden Evaluation steht an erster Stelle der Vergleich von Kindern derselben nicht-deutschen Muttersprache in bilingualen versus monolingualen Klassen, um die Frage nach der besseren Wirkung eines Unterrichtsmodells auf die Deutschkenntnisse und den Schulerfolg beantworten zu können. Wenn darüber hinaus monolingual deutsche Kinder in die Experimental- und Kontrollgruppe aufgenommen und dieselben Tests und Hintergrundvariablen erhoben würden (ohne sie in den Analysen mit den bilingualen zu vermischen), ließe sich zudem die Frage beantworten, wie die Nicht-Muttersprachler im Zeitverlauf und unter statistischer Berücksichtigung wesentlicher Einflussfaktoren gegenüber Muttersprachlern abschneiden, bzw. ob sie idealerweise „aufgeholt“ haben.

---

<sup>25</sup> vgl. Darstellungen zu bilinguale Grundschulen für unterschiedliche Partnersprachen unter: <http://fhh.hamburg.de/stadt/Aktuell/hamburgservice/broschueren/start.html>

<sup>26</sup> Ähnlich der international vergleichenden Makroanalysen, in denen Staaten die Analyseeinheit sind, wäre also auch bei bilingualen Schulen in Deutschland eine Mehrebenenanalyse „hampered by the low number of macro cases“ (van Tubergen u.a. 2004: 706).



## 7. Resümee

Die Ausgangsfrage dieses Berichts war, ob für Kinder mit Minderheiten- oder Migrantensprachen als Erstsprache (L1) zweisprachiger Unterricht effektiver für den Erwerb der Zweitsprache und die Steigerung Schulleistungen ist als einsprachige Unterrichtsformen. Nach der detaillierten Auswertung sowohl der Metaanalysen von Greene und Slavin/Cheung als auch ausgewählter Originalstudien lässt sich dies vorläufig wie folgt beantworten:

In den überprüften methodisch besseren Studien gibt es keine Anzeichen dafür, dass sich zweisprachige Schulprogramme negativ auf die schulischen Leistungen in der Zweitsprache (L2) auswirken. Dies wurde insbesondere bezüglich des Einflusses auf die Lesekompetenz untersucht. Diese relativ sichere Grundaussage, dass bilingualer Unterricht bzw. die zusätzliche Unterrichtung der und in der Muttersprache nicht schadet, impliziert auch, dass die „time-on-task-Hypothese“ (s. oben, S. 11) in diesem Kontext nicht bestätigt werden konnte: Wenn ein Teil der Unterrichtsstunden in einer anderen Sprache als der des Regelunterrichts erteilt wird, führt das offenbar nicht automatisch zu schlechteren Leistungen in der Zweitsprache und den Sachfächern. Ein in den Studien nicht systematisch untersuchter Grund dafür mag sein, dass bei diesen bilingual unterrichteten Kindern auch außerhalb der Klasse, in der Schule, der Nachbarschaft, mit Freunden und in den Medien viel Kommunikation in L2 vonstatten geht und so die entgangenen Schulstunden in L2 „kompensiert“ werden.

Für die Hypothese, dass sich bilinguale Programme bzw. muttersprachlicher Unterricht nicht nur neutral, sondern positiv auf die L2-Kompetenz und Schulleistungen in L2 auswirken, liegen beim derzeitigen Forschungsstand weder eindeutige noch verlässliche Nachweise vor. Je nach Unterrichtsmodell und weiteren Kontextmerkmalen variiert der Effekt zwischen neutral und teilweise signifikant positiv. Die meisten Evaluationsstudien haben allerdings so gravierende Schwächen, dass diese Aussagen nur unter großem Vorbehalt getroffen werden können (mit einer ähnlich kritischen Einschätzung der internationalen Forschung und ihrer einseitigen Rezeption auch in Deutschland vgl. Hopf 2005: 241).

Insbesondere einige Publikationen, die sich unumwunden für bilingualen Unterricht aussprechen und die häufig als Beleg für die positive Wirkung bilingualer Erziehung angeführt werden, haben sich als wissenschaftlich inakzeptabel herausgestellt: Dies sind allen voran die Studie von Thomas und Collier (1997; 2002) sowie die Metaanalyse von Greene (1998). Auch die Ramirez-Studie (1991) kann nicht als Argument zugunsten eines bilingualen Unterrichts herangezogen werden, sondern bietet den methodisch überzeugenden Nachweis, dass zumindest die untersuchte Variante von transitorischen bilingualen (early-exit) Programmen Kindern bis zur dritten Klasse nicht mehr Unterstützung bietet als monolingualer Unterricht. Allerdings wurde dieses Patt im Vergleich zu einem einsprachigen Förderunterricht festgestellt, der den Spracherwerb durch systematisches Lehren von Englisch als Zweitsprache unterstützt, und nicht zu der so genannten „sink or swim“-Alternative des reinen monolingualen Regelunterrichts.

Legt man die vorliegende Evidenz vorsichtig dahingehend aus, dass bilinguale Erziehung den Kindern mit Minderheitensprachen als L1 nicht schadet (sie gleichzeitig ihnen aber auch nicht durchgängig hilft), so spricht dies gegen die Allgemeingültigkeit der Schwellenhypothese von Cummins. Danach muss erst in L1 eine nötige „akademische“ Kompetenz, also eine Alphabetisierung, erreicht werden, damit ähnliche L2-Kompetenzen – erleichtert durch metalinguistische Transfers (Interdependenz-Hypothese) – erlangt werden können. Zumindest wenn ein Programm wie die transitorischen (early-exit) Modelle diese unterstellte Sprachentwicklung „nachbaut“, scheint dies nicht per se förderlich zu sein. Dies kann daran liegen, dass in Cummins Thesen zu hohe Erwartungen an die Wirksamkeit der sprachlichen Prozesse formuliert werden, wogegen Faktoren unterschätzt werden, die das erreichte Niveau in L1 und L2 und die kognitive Entwicklung allgemein beeinflussen (z.B. Intelligenz, familiärer Bildungshintergrund, Gelegenheiten zum Sprachlernen und zur Anwendung, Motivation). Insbesondere scheint bilinguale Unterrichtung an sich nicht das Allheilmittel für auch familiär bedingte schlechtere Lernausgangslagen etwa aufgrund eines bildungsfernen Elternhauses zu sein. So argumentiert etwa Rösch in Bezug auf das Berliner Modell zweisprachiger Alphabetisierung türkischer SchülerInnen, es fehle zwar der Nachweis, „dass die zweisprachige Erziehung den Zweitspracherwerb und den Schulerfolg türkischer Kinder deutlich verbessert. Dass sie [die wissenschaftliche Begleitung; Anm d. V.] dies versucht hat und vielleicht [...] auch nur unter dieser Prämisse eingesetzt worden ist, zeigt das Dilemma, in dem sich Spracherziehungskonzepte für Minderheitenangehörige immer wieder befinden: Mit ihnen sollten die gesamten Probleme, die die Schule mit dieser Schülerschaft und die die Schüler mit dieser Schule haben, gelöst werden. Sprache und Zweisprachigkeit wird dabei so sehr funktionalisiert, dass der Misserfolg vorprogrammiert ist.“ (Rösch 2001: 33)

Die eigentliche Leistung von Modellen zweisprachiger Erziehung besteht darin, dass Kinder zusätzlich in ihrer Muttersprache lesen und schreiben lernen und dabei offenbar mit monolingual unterrichteten Kindern derselben Herkunftssprache in den (in L2 gemessenen) Schulleistungen zumindest gleichauf liegen, obwohl sie (je nach Modell) weniger Unterrichtsstunden in L2 haben. Dieser Zusatzkompetenz der L1-Alphabetisierung, die Kinder im Regelunterricht nicht erwerben, kann selbstverständlich ein hoher Wert zugesprochen werden.

Ob ein spezifisches bilinguales Unterrichtsmodell gegenüber einem monolingualen effektiver in der Vermittlung von L2 und fachlichen Kompetenzen ist, wird nicht zuletzt von der Qualität des Unterrichts abhängen. Unbestritten ist, dass eine durchdachte didaktische Konzeption und pädagogisch hoch qualifiziertes Lehrpersonal den eigentlichen Ausschlag geben können. Dies zeigte sich etwa an den exemplarischen Schulmodellen von Campeau u.a. (1975) bzw. an dem diesbezüglich weniger erfolgreichen muttersprachlichen Unterricht in den Niederlanden.

Aus deutscher Perspektive sind vor allem US-Studien von Interesse, die two-way-Programme mit koordinierter Alphabetisierung in L1 und L2 evaluieren, da diese am ehesten den deutschen Modellen bilingualer Grundschulen mit Migrantensprachen als Partnersprachen entsprechen. Die Frage nach der Effektivität dieser Modelle ist für Deutschland noch völlig offen, da bisher keine Evaluationsstudie oder wissenschaftliche Begleitfor-

schung von Modellklassen vorliegt, die aufgrund ihres Forschungsdesigns diese Fragestellung beantworten könnte. Methodisch entscheidend wäre in einer neuen Studie der Vergleich der sprachlichen oder schulischen Kompetenzen mit einer Kontrollgruppe von Kindern der gleichen Herkunftssprache im Regelunterricht bzw. optimalerweise in einem alternativen deutschsprachigen Förderunterricht mit innovativen Deutsch-als-Zweitsprache-Methoden.

Mit einer neuen Studie, für die in Kapitel 6 einige weitere Grundüberlegungen angestellt wurden, ließe sich auch eine Forschungslücke schließen, die bei den US-amerikanischen Studien offen bleibt: Dort wird die Wirkung bi- und monolingualer Programme nicht wirklich differenziert genug dahingehend untersucht, welche Englischkenntnisse die Kinder zur Einschulung mitbringen (zu unterschiedlichen Zielgruppen vgl. auch Glenn 1996: 403). Die übliche Einstufung als „limited English proficient“ ist zu grobschlächtig. Denn aus der Forschung zum Zweitspracherwerb ist bekannt, dass es sehr unterschiedliche Ausprägungen von Bilingualismus gibt. Es „ist mit erheblichen Variationen des Verhältnisses von Erst- und Zweitsprache von Individuum zu Individuum und im Verlauf des Lern- und Entwicklungsprozesses jedes Einzelnen zu rechnen, die je unterschiedliche Förderentscheidungen nahe legen.“ (Ehlich u.a. 2004: 16; vgl. hierzu auch Preibusch 1992) Wirken Programme unterschiedlich, wenn die Kinder etwa, da sie neu zugewandert sind, gar keine L2-Kenntnisse haben oder wenn in verschiedenen linguistischen Teilbereichen Stärken oder Schwächen in L2-Kompetenzen festzustellen sind? Hier bleibt der Einsatz von Verfahren zur Feststellung des Sprachstands unersetzlich, wie sie in der Bundesrepublik derzeit entwickelt werden. Auch eine ausgefeiltere Berücksichtigung der Migrationsbiografie der Kinder bzw. ihrer Eltern wäre ein Fortschritt gegenüber den meisten der US-Studien, die im Rahmen ihres Konzepts von Sprachminderheiten den Zuwanderungsaspekt häufig vernachlässigen.

Weitgehend offen bleibt also die Frage nach Interaktionseffekten zwischen der Wirkung bilingualer Programme bei Schülergruppen mit unterschiedlicher Migrationsbiografie und unterschiedlichen L1- und L2-Kenntnissen. Auch Wechselwirkungen mit Kontextmerkmalen wie der durchschnittlichen Schichtzugehörigkeit und dem Grad der ethnisch-linguistischen Segregation im Klassenverband oder weiter gefasst im Wohnviertel gilt es noch systematischer zu untersuchen. Ebenso selten wird in den US-amerikanischen Studien thematisiert, wie groß der Effekt der zweisprachigen Unterrichtung auf die Leistungen relativ zum Einfluss anderer gewichtiger Einflussfaktoren ist. Ramirez u.a. (1991) stellen teilweise leicht positive Wirkungen der elterlichen Verbundenheit zur Schriftkultur auf die Leseleistung fest. Da aber hier wie bei vielen anderen Studien die meisten der teilnehmenden SchülerInnen aus sozial schwächeren Familien kamen und Schulen mit hohen Anteilen spanischsprechender SchülerInnen besuchten, lässt sich die Frage nach Interaktionseffekten bzw. nach relativen Effekten bisher nicht gründlich klären: Wie würde die Maßnahme in einer Umgebung wirken, wenn nur eine Minderheit aller Kinder die Verkehrssprache nicht als L1 spricht? Haben bilinguale SchülerInnen aus höheren Schichten größere Vorteile vom Besuch eines bilingualen Programms als solche aus sozio-ökonomisch benachteiligten Familien? Mit solchen Erkenntnissen ließe sich realistischer einschätzen, wie groß die Einflussmöglichkeit schulischer Intervention mittels eines anspruchsvoll konzipierten und

durchgeführten bilingualen Unterrichts ist – gerade angesichts des starken Zusammenhangs von familiärem Hintergrund und schulischen Leistungen.

Systematische Evaluationsstudien zu bilingualen Modellen im deutschen Kontext könnten so auch die wissenschaftliche Erkenntnisgewinnung vorantreiben, die eben auch in dem „Evaluierungsland“ USA keineswegs perfekt ist. Die Anwendung systematischer Evaluationsmethoden ist für wissenschaftliche und bildungspolitische Gegner wie Befürworter zweisprachiger Programme wichtig. Denn bei offensichtlichen methodischen Defiziten kann die Kritikerin vermeintliche Erfolge zu Recht in Zweifel ziehen oder der Befürworter die vielleicht tatsächlichen Erfolge nicht wasserdicht nachweisen – und umgekehrt.

Bei der Durchsicht der vorhandenen Forschung hat sich herausgestellt, dass es sich hier um eine Fragestellung handelt, die viel stärker als bisher in interdisziplinären Forscherteams bearbeitet werden sollte. Andernfalls neigen Forschungsberichte dazu, auf dem einen oder anderen methodischen Auge blind zu sein. Die quantitativ empirisch arbeitende Soziologie oder Psychologie könnten insbesondere ihren Sachverstand zu der Konzeption von Quasi-Experimenten und multivariaten statistischen Analysen einbringen. Die Linguistik bzw. Germanistik ist wiederum unverzichtbar für das Verständnis erst- und zweitsprachlicher Aneignungsprozesse – etwa auch in Bezug auf die Transfers zwischen Sprachen, die unterschiedlich eng verwandt sind – sowie für die Entwicklung von Sprachtests als den bedeutendsten Messinstrumenten in solchen Evaluationsstudien. Die Erziehungswissenschaft schließlich sollte ihre besondere Kompetenz bezüglich der Konzeptionierung der Unterrichtsmodelle und didaktischen Methoden sowie bei deren Implementierung in den Schulen einbringen. Die Stärken der unterschiedlichen Disziplinen in einem zukünftigen Evaluierungsprojekt zu bündeln, wäre also viel versprechend.

Insgesamt wäre in der wissenschaftlichen Forschung eine produktive Konkurrenz bei der Konzeption und Untersuchung bi- und monolingualer Förderprogramme wünschenswert. Denn einerseits werden schulische Programme, Unterrichtsmethoden und Lehrpläne für Deutsch als Zweitsprache von vielen Kultusministerien zunehmend gefördert (Zickgraf 2005). Gleichzeitig wird in der gesellschaftlichen Diskussion die Förderung von Mehrsprachigkeit immer wieder eingefordert, auch mit Blick auf ihren möglichen individuell beruflichen und volkswirtschaftlichen Nutzen. Eine vergleichende Auswertung solcher Ansätze wäre also zu begrüßen. Letztlich wären eine Versachlichung der Diskussion und eine Förderung von unterschiedlichen Programmen wünschenswert, die sich, wissenschaftlich belegt, für verschiedene Schülergruppen als erfolgreich erwiesen haben. Unzweifelhaft sollten didaktische Interventionen, die die Deutschkenntnisse oder zusätzlich eine kompetente Zweisprachigkeit fördern und zudem soziale Benachteiligung reduzieren wollen, nicht erst mit der Schulpflicht ab der ersten Klasse einsetzen, sondern bereits früh im Kindergarten ansetzen (zu bisherigen Ansätzen elementarpädagogischer Sprachförderung in Deutschland vgl. Jampert u.a. 2005). Linguistische Erkenntnisse zum Spracherwerb favorisieren eine solche frühere Option, die es gerade für die hier geborenen Kinder aus Zuwandererfamilien wesentlich wahrscheinlicher macht, dass sie mit Deutschkenntnissen auf muttersprachlichem Niveau ihre Schulkarriere starten.

Bilinguale Unterrichtsmodelle, die in allen Klassen oder zumindest während der Grundschule Sprach- und Sachunterricht in zwei Sprachen erteilen, sind mit erheblichen finanziel-

len Investitionen verbunden: Es werden kompetente bilinguale Lehrkräfte gebraucht, die an Universitäten dazu ausgebildet werden, Fächer aller Jahrgangsstufen in einer anderen als der Landessprache, aber gemäß der bundesländerspezifischen Lehrpläne und pädagogisch-didaktischen Standards zu unterrichten. Solch eine zusätzliche oder parallele Lehrerausbildung und Curriculumentwicklung wäre ein nicht zu vernachlässigender Kostenpunkt, insbesondere wenn man zweisprachigen Unterricht für mehrere Migrantensprachen anbieten wollte. Solange man sich nicht mit dem Status Quo an deutschen Schulen und ihrer mangelhaften Kompetenzvermittlung an MigrantenschülerInnen zufrieden geben will, wären allerdings auch alternative Fördermodelle, die weitgehend mit der Unterrichtssprache Deutsch arbeiten (z.B. Deutsch als Zweitsprache, Einstellung von mehr Lehrkräften zur Verkleinerung von Klassen und Teamteaching innerhalb von Klassen etc.), nicht umsonst zu haben.

## Literatur

- Alanís, Iliana (2000): A Texas two-way bilingual program: Its effects on linguistic and academic achievement. In: *Bilingual Research Journal*, 24 (3): 225-248.
- Allemann-Ghionda, Cristina (1999): Schule, Bildung und Pluralität. Sechs Fallstudien im europäischen Vergleich. Bern u.a.: Peter Lang.
- Alvarez, J. (1975): Comparison of academic aspirations and achievement in bilingual versus monolingual classrooms. Austin, TX, University of Texas at Austin.
- Axelsson, Monica (2005): Mother tongue teaching and programs for bilingual children in Sweden. In: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI) (Hrsg.). The effectiveness of bilingual school programs for immigrant children. Berlin: Wissenschaftszentrum Berlin für Sozialforschung, Discussion paper SP IV 2005-601, 108-122.
- Bacon, Herbert L., Gerald D. Kidd und John J. Seaberg (1982): The effectiveness of bilingual instruction with Cherokee Indian students. In: *Journal of American Indian Education*, 21 (1): 34-43.
- Baker, Colin (2001): Foundations of bilingual education and bilingualism. Clevedon: Multilingual Matters Ltd.
- Baker, Keith (1992): Ramirez et al. misled by bad theory. In: *Bilingual Research Journal*, 16 (1&2): 157-169.
- Baker, Keith und A. deKanter (1981): Effectiveness of bilingual education: A review of the literature (Final draft report). Washington: Office of Technical and Analytic Systems, U.S. Department of Education.
- Baumann, Till (2001): Mehrsprachigkeit an deutschen Schulen. Ein Länderüberblick. Berlin/Bonn: Die Beauftragte der Bundesregierung für Ausländerfragen, In der Diskussion Nr. 10.
- Baur, Rupprecht S. (2000): Schulischer Zweitspracherwerb bei Migrantenschülern - Theorie und Empirie. In: *Deutsch als Fremdsprache*, 37 (2): 131-136.
- Beck, Eric (1999): Language rights and Turkish children in Germany. In: *Patterns of Prejudice*, 33 (2): 3-12.
- Beelmann, Andreas und Thomas Bliesener (1994): Aktuelle Probleme und Strategien der Metaanalyse. In: *Psychologische Rundschau*, 45: 211-233.
- Bialystok, Ellen (2004): The impact of bilingualism on language and literacy development. In: Tej K. Bhatia und William C. Ritchie (Hrsg.). *The handbook of bilingualism*. Malden/MA: Blackwell, 577-601.
- Birdsong, David (Hrsg.) (1999): *Second language acquisition and the critical period hypothesis*. Mahwah & London: Lawrence Erlbaum.
- BLK (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung) (2004): BLK-Programm Förderung von Kindern und Jugendlichen mit Migrationshintergrund FÖRMIG. Bonn, <http://www2.erzwiss.uni-hamburg.de/aktuell/foermig.pdf>.
- BMFSFJ (Bundesministerium für Familie, Senioren, Frauen und Jugend) (Hrsg.) (2004): *Viele Welten leben. Lebenslagen von Mädchen und jungen Frauen mit griechischem, italienischem, jugoslawischem, türkischem und Aussiedlerhintergrund*. Berlin.
- Bortz, Jürgen (1999): *Statistik für Sozialwissenschaftler*. Berlin/Heidelberg: Springer.
- Bos, Wilfried, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Gerd Walther und Renate Valtin (Hrsg.) (2003): *Erste Ergebnisse aus IGLU*. Münster: Waxmann Verlag.

- Broeder, Peter und Guus Extra (1999): *Language, ethnicity and education. Case studies on immigrant minority groups and immigrant minority languages*. Clevedon u.a.: Multilingual Matters Ltd.
- Brumlik, Micha (2000): Wann ist muttersprachlicher Unterricht sinnvoll? In: *Erziehung & Wissenschaft* (April): 21.
- Büchel, Helga und Sabine Bühler-Otten (2000): Förderung von Zweisprachigkeit: Herkunftssprachenunterricht in Hamburg. In: *Deutsch Lernen*, 25 (2): 179-186.
- Bundesforum Familie (2004): *Grundsätze und Perspektiven einer familienorientierten Integrationspolitik - Migrationsfamilien zwischen Integration und Ausgrenzung*. Berlin.
- Butler, Yuko G. und Kenji Hakuta (2004): Bilingualism and second language acquisition. In: Tej K. Bhatia und William C. Ritchie (Hrsg.). *The handbook of bilingualism*. Malden/MA: Blackwell, 114-144.
- Campeau, Peggie L., A. Oscar H. Roberts, John E. Bowers, Melanie Austin und Sarah J. Roberts (1975): *The identification and description of exemplary bilingual education programs*. Palo Alto, CA: American Institutes for Research.
- Carlisle, Joanne F. und F.F. Beeman (2000): The effects of language of instruction on the reading and writing achievement of first-grade Hispanic children. In: *Review of Educational Research*, 55 (3): 269-317.
- Carroll, John Bissell (1963): A model of school learning. In: *Teachers' College Record*, 64 (8): 723-733.
- Cohen, Andrew D. (1975): *A sociolinguistic approach to bilingual education*. Rowley, MA: Newbury House Press.
- Cook, Thomas D. und Donald T. Campbell (1979): *Quasi-experimentation. Design & analysis issues for field settings*. Chicago: Rand McNally College Publishing Company.
- Cooper, Harris und Larry V. Hedges (Hrsg.) (1994): *The handbook of research synthesis*. New York: Russel Sage Foundation.
- Covey, Donald David (1973): An analytical study of secondary freshman bilingual education and its effect on academic achievement and attitude of Mexican American students, Arizona State University.
- Cummins, Jim (1979): Linguistic interdependence and the educational development of bilingual children. In: *Review of Educational Research*, 49 (2): 222-251.
- Cummins, Jim (1981): The role of primary language development in promoting educational success for language minority students. In: California State Department of Education (Hrsg.). *Schooling and language minority students: a theoretical framework*. Los Angeles: Evaluation, Dissemination and Assessment Center, California State University, 3-49.
- Cummins, Jim (1982): Die Schwellenniveau- und Interdependenz-Hypothese: Erklärungen zum Erfolg zweisprachiger Erziehung. In: James Swift (Hrsg.). *Bilinguale und multikulturelle Erziehung*. Würzburg: Königshausen + Neumann, 34-43.
- Cummins, Jim (1986): Language proficiency and academic achievement. In: Jim Cummins und Merrill Swain (Hrsg.). *Bilingualism in education: aspects of theory, research, and practice*. London/New York: Longman, 138-161.
- Cummins, Jim (2003): Bilingual Education. In: Jill Bourne und Euan Reid (Hrsg.). *World Yearbook of Education*. London: Kogan Page, 3-19.

- Danoff, Malcom N. (1978): Evaluation of the impact of ESEA Title VII Spanish/English bilingual education program: American Institutes for Research.
- de Swaan, Abram (2001): Words of the World: The Global Language System. Cambridge: Polity.
- Deutscher Bildungsserver (2004): Wege, Versuche, Erlasse zum bilingualen Unterricht, <http://lernen.bildung.hessen.de/bilingual/bildungspolitik/versuche> [15.12.2004].
- Deutsches PISA-Konsortium (Hrsg.) (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.
- Driessen, Geert (2005): From cure to curse: The rise and fall of bilingual education programs in the Netherlands. In: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI) (Hrsg.). The effectiveness of bilingual school programs for immigrant children. Berlin: Wissenschaftszentrum Berlin für Sozialforschung, Discussion Paper SP IV 2005-601, 77-108.
- Driessen, Geert, Kees De Bot und Paul Jungbluth (1989): De effectiviteit van het Onderwijs in Eigen Taal en Cultuur. Prestaties van Marokkaanse, Spaanse en Turkse leerlingen. Nijmegen: ITS.
- Edwards, John V. (2004): Foundations of bilingualism. In: Tej K. Bhatia und William C. Ritchie (Hrsg.). The handbook of bilingualism. Malden/MA: Blackwell, 7-31.
- Ehlich, Konrad, Huub van den Bergh, Ursula Bredel, Birgitta Garne, Anna Komor, Hans-Jürgen Krumm, Tim McNamara, Hans H. Reich, Guido Schnieders und Jan D. ten Thije (2004): Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Sprachförderung von Kindern mit und ohne Migrationshintergrund. Eine Expertise. München.
- Felix, Sascha (1993): Psycholinguistische Untersuchung zur zweisprachigen Alphabetisierung. Gutachten im Auftrag der Berliner Senatsverwaltung für Schule. Berlin.
- Fillia, Elena (2003): Das bilinguale deutsch-italienische Schulprojekt in Frankfurt am Main. In: Jürgen Erfurt, Gabriele Budach und Sabine Hofmann (Hrsg.). Mehrsprachigkeit und Migration. Frankfurt a.M.: Peter Lang, 212-226.
- Fitz-Gibbon, Carol Taylor und Lynn Lyons Morris (1978): How to design a program evaluation. Beverly Hills/London: Sage Publications.
- Frey, Sigfried und Hans-Goerg Frenz (1982): Experiment und Quasi-Experiment im Feld. In: Jean-Luc Patry (Hrsg.). Feldforschung: Methoden und Probleme sozialwissenschaftlicher Forschung unter natürlichen Bedingungen. Bern: Huber.
- Fried, Lilian (2004): Expertise zu Sprachstandserhebungen für Kindergartenkinder und Schulanfänger. Eine kritische Betrachtung. München: Deutsches Jugendinstitut.
- Fürstenau, Sara, Ingrid Gogolin und Kutlay Yamgur (Hrsg.) (2003): Mehrsprachigkeit in Hamburg. Ergebnisse einer Sprachenerhebung an den Grundschulen in Hamburg. Münster/New York/München/Berlin: Waxmann Verlag.
- Garcia, Eugene E. (2002): Bilingualism and schooling in the United States. In: International Journal of the Sociology of Language, Nr. 155-156: 1-92.
- Glenn, Charles (1996): Educating immigrant children. Schools and language minorities in twelve nations. New York/London: Garland Publishing, Inc.
- Gogolin, Ingrid (2000): Bildung und ausländische Familien. In: Sachverständigenkommission 6. Familienbericht (Hrsg.). Familien ausländischer Herkunft in Deutschland: Lebensalltag. Opladen: Leske + Budrich, 61-106.



- Gogolin, Ingrid (2005): Bilingual education - The German experience and debate. In: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI) (Hrsg.). The effectiveness of bilingual school programs for immigrant children. Berlin: Wissenschaftszentrum Berlin für Sozialforschung, Discussion Paper SP IV 2005-601, 133-145.
- Gogolin, Ingrid, Ursula Neumann und Lutz Reuter (Hrsg.) (2001a): Schulbildung für Kinder aus Minderheiten in Deutschland 1989-1999. Münster/ New York/München/Berlin.
- Gogolin, Ingrid, Ursula Neumann und Hans-Joachim Roth (2001b): Anforderungen an Verfahren zur Erhebung sprachlicher Fähigkeiten und des Sprachgebrauchs zweisprachiger Kinder. Stellungnahme zur "Berliner Sprachstandserhebung und Materialien zur Sprachförderung für Kinder in der Schuleingangsphase", <http://www.erzwiss.uni-hamburg.de/Arbeitsstellen/Interkultur/baere.html>.
- Gogolin, Ingrid, Ursula Neumann und Hans-Joachim Roth (2003): Förderung von Kindern und Jugendlichen mit Migrationshintergrund. Bonn: Bund-Länder-Kommission (BLK), Materialien zur Bildungsplanung und Forschungsförderung, Heft 107.
- Gogolin, Ingrid und Ludger Pries (2004): Stichwort: Transmigration und Bildung. In: Zeitschrift für Erziehungswissenschaft, 7 (1): 5-19.
- Gräfe-Bentzien, Sigrid (2001): Evaluierung bilingualer Kompetenz. Eine Pilotstudie zur Entwicklung der deutschen und italienischen Sprachfähigkeiten in der Primarstufe beim Schulversuch der Staatlichen Europa-Schule Berlin (SESB). Berlin: Freie Universität Berlin, <http://www.diss.fu-berlin.de/2001/14/index.html>.
- Greene, Jay P. (1997): A meta-analysis of the Rossell and Baker review of bilingual education research. In: Bilingual Research Journal, 21 (2&3): 103-122.
- Greene, Jay P. (1998): A meta-analysis of the effectiveness of bilingual education. Los Angeles: University of Southern California, Tomas Rivera Policy Institute, <http://ourworld.compuserve.com/homepages/JWCRAWFORD/greene.htm>.
- Grießhaber, Wilhelm (2001): Verfahren und Tendenzen der funktional-pragmatischen Diskursanalyse. Vom Speiserestaurant zum Cybercafé. In: Zsuzsanna Ivanyi und Andras Kertész (Hrsg.). Gesprächsforschung: Tendenzen und Perspektiven. Frankfurt/M. u.a: Lang, 75-95.
- Grießhaber, Wilhelm (2004): Zu Cummins (1980): BICS vs. CALP, <http://spzwww.uni-muenster.de/~griesha/sla/cummins/eisberg.html> [15.12.2004].
- Grin, François, Jean Rossiaud und Bülent Kaya (2003): Immigrationssprachen und berufliche Integration in der Schweiz. In: Hans-Rudolf Wicker, Rosita Fibbi und Werner Haug (Hrsg.). Migration und die Schweiz. Zürich: Seismo-Verlag, 421-452.
- Haase, Gisela (2003): Migration und Mehrsprachigkeit in der schulischen Praxis. Die Georg Büchner-Schule in Frankfurt-Bockenheim. In: Jürgen Erfurt, Gabriele Budach und Sabine Hofmann (Hrsg.). Mehrsprachigkeit und Migration. Frankfurt a.M. u.a.: Peter Lang, 217-220.
- Heckman, James J. (1979): Sample Selection Bias as a Specification Error. In: Econometrica, 47 (1): 153-161.
- Heintze, Andreas (2001): Erfolgsstory oder Sackgasse. Zur Kontroverse um zweisprachige Erziehung in den USA. In: Deutsches Jugendinstitut - Projekt "Kulturenvielfalt" (Hrsg.). Treffpunkt deutsche Sprache. Sprachförderung von mehrsprachigen Kindern in Tageseinrichtungen. München: Deutsches Jugendinstitut (DJI), 33-47.
- Hill, Margret (1995): Invandrarbarns möjligheter. Om kunskapsutveckling och språkutveckling i förskola och skola. Göteborg: Göteborgs universitet, Institutionen för pedagogik.

- Hofstetter, Carolyn Huie (2004): Effects of a transitional bilingual education program: Findings, Issues, and Next Steps. In: *Bilingual Research Journal*, 28 (3): 355-377.
- Hopf, Diether (2005): Zweisprachigkeit und Schulleistung bei Migrantenkindern. In: *Zeitschrift für Pädagogik*, 51 (2): 236-251.
- Hunger, Uwe und Dietrich Thränhardt (2004): Migration und Bildungserfolg: Wo stehen wir? In: *IMIS-Beiträge*, Nr. 23: 179-197.
- Huzar, Helen (1973): The effects of an English-Spanish primary grade reading program on second and third grade students, Rutgers University.
- Imhoff, Gary (1990): The position of U.S. English on bilingual education. In: *The Annals of the American Academy of Political and Social Science*, Nr. 508 (April): 48-61.
- Jacobs, Bernhard (1998): Interne Validität,  
<http://www.phil.uni-sb.de/~jacobs/seminar/vpl/validitaet/interne.htm#Begriff>.
- Jampert, Karin, Petra Best, Angela Guadatiello, Doris Holler und Anne Zehnauer (2005): Schlüsselkompetenz Sprache. Sprachliche Bildung und Förderung im Kindergarten. Konzepte, Projekte und Maßnahmen. Weimar/Berlin: Das Netz.
- Kaufman, Maurice (1968): Will instruction in reading Spanish affect ability in reading English. In: *Journal of Reading*, 11: 521-527.
- Kleff, Sanem (1998): Modelle Bilingualer Erziehung in Berlin. In: *Landesarbeitsgemeinschaft der Ausländerbeiräte Nordrhein-Westfalen* (Hrsg.). *Aus vielen Quellen schöpfen: Mehrsprachigkeit, religiöse und kulturelle Identität; eine Bereicherung für diese Gesellschaft*. Düsseldorf, 65-77.
- Knapp, Werner (1999): Verdeckte Sprachschwierigkeiten. In: *Die Grundschule*, 5 (99): 30-33.
- Kromrey, Helmut (1998): *Empirische Sozialforschung*. 8. Auflage. Opladen: Leske + Budrich.
- Kuhs, Katharina und Wolfgang Steinig (Hrsg.) (1998): *Pfade durch Babylon. Konzepte und Beispiele für den Umgang mit sprachlicher Vielfalt in Schule und Gesellschaft*. Freiburg: Fillibach Verlag.
- Kultusministerkonferenz (Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (2002): Bericht "Zuwanderung". Beschluss der Kultusministerkonferenz vom 24.05.2002. Bonn.
- Kupfer-Schreiner, Claudia (1994): *Sprachdidaktik und Sprachentwicklung im Rahmen interkultureller Erziehung. Das Nürnberger Modell. Ein Beitrag gegen Rassismus und Ausländerfeindlichkeit*. Weinheim: Deutscher Studien Verlag.
- Lambert, Wallace E. und G. Richards Tucker (1972): *Bilingual education of children: The St. Lambert project*. Rowley, MA: Newbury House.
- Lindholm-Leary, Kathryn J. (2001): *Dual language education*. Clevedon/Buffalo/ oronto/Sydney: Multilingual Matters Ltd.
- López, María G. und Abbas Tashakkori (2004): Effects of a two-way bilingual program on the literacy development of students in kindergarten and first grade. In: *Bilingual Research Journal*, 28 (1): 19-34.
- Maas, Utz und Ulrich Mehlem (2003): *Qualitätsanforderungen für die Sprachförderung im Rahmen der Integration von Zuwanderern*. Osnabrück: IMIS.

- Maas, Utz, Ulrich Mehlem und Christoph Schroeder (2004): Mehrsprachigkeit und Mehrschriftigkeit bei Einwanderern in Deutschland. In: Klaus J. Bade, Michael Bommes und Rainer Münz (Hrsg.). Migrationsreport 2004. Frankfurt/ New York: Campus Verlag, 117-150.
- Maldonado, J.R. (1977): The effect of the ESEA Title VII program on the cognitive development of Mexican American students. Houston, TX, University of Houston.
- Maldonado, J.R. (1994): Bilingual special education: Specific learning disabilities in language and reading. In: Journal of Education Issues of Language Minority Students, 14 (Winter): 127-147.
- May, Anna (2003): Zweisprachigkeit durch zweisprachige Erziehung? Die Staatliche Europa-Schule Berlin - ein Konzept und seine Realisierung: Osnabrück.
- Meisel, Jürgen M. (2003): Zur Entwicklung der kindlichen Mehrsprachigkeit: Unveröfftl. Gutachten für die Bildungskommission der Herinrich-Böll-Stiftung.
- Meisel, Jürgen M. (2004): The bilingual child. In: Tej K. Bhatia und William C. Ritchie (Hrsg.). The handbook of bilingualism. Malden/MA: Blackwell, 93-113.
- Meyer, Michael M. und Stephen E. Feinberg (1992): Assessing evaluation studies: The case of bilingual education strategies. Washington: National Academy Press.
- Moran, Carrol E. und Kenji Hakuta (2001): Bilingual education: Broadening research perspectives. In: James A. Banks (Hrsg.). Handbook of research on multicultural education. San Francisco: Jossey-Bass, 445-462.
- Mouw, Ted und Yu Xie (1999): Bilingualism and the academic achievement of first- and second-generation Asian Americans: Accommodation with or without assimilation? In: American Sociological Review, 64 (2): 232-252.
- Nehr, Monika, Karin Birnkott-Rixius, Leyla Kubat und Sigrid Masuch (1988): In zwei Sprachen lesen lernen - geht denn das? Erfahrungsbericht über die zweisprachige koordinierte Alphabetisierung. Weinheim/Basel: Beltz Verlag.
- Nehr, Monika und Edeltraut Karajoli (1995): Expertise on bilingual literacy education of Turkish schoolchildren in Berlin. Berlin: Council of Europe.
- Nitzschke, Volker (Hrsg.) (1982): Multikulturelle Gesellschaft - multikulturelle Erziehung? Stuttgart: Metzler.
- OECD (2002): Sample tasks from the PISA 2000 assessment. Reading, mathematical and scientific literacy, OECD.
- Parliamentary Inquiry Committee on Integration Policy (2004): Summary, conclusions and recommendations: Dutch House of Representatives, <http://www.dutchuec.nl/kcgs/show/id=136103/contentid=5111> [3.2.2005.2004].
- Plante, Alexander J. (1978): A study of effectiveness of the Connecticut "paring" model of bilingual/bicultural education. Hamden, Connecticut Staff Development Cooperative.
- Portes, Alejandro und Lingxin Hao (2002): The price of uniformity: language, family and personality adjustment in the immigrant second generation. In: Ethnic and Racial Studies, 25 (6): 889-912.
- Portes, Alejandro und Rubén G. Rumbaut (2001): Legacies. The story of the immigrant second generation. Berkley/Los Angeles/London: University of California Press.
- Powers, Stephen (1978): The influence of bilingual instruction on academic achievement and self-esteem of selected Mexican-American junior high school students, The University of Arizona.

- Preibusch, Wolfgang (1992): Die deutsch-türkischen Sprachbalancen bei türkischen Grundschulern. Eine clusteranalytische Untersuchung. Frankfurt a.M.: Lang.
- Ramirez, J. David, Sandra D. Yuen und Dena R. Ramey (1991a): Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children. Final report - Executive summary. San Mateo: Aguirre International.
- Ramirez, J. David, Sandra D. Yuen, Dena R. Ramey und David J. Pasta (1991b): Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children. Final report - Volume I. San Mateo: Aguirre International.
- Ramirez, J. David, David J. Pasta, Sandra D. Yuen, David K. Billings und Dena R. Ramey (1991c): Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children. Volume II - Final report. San Mateo: Aguirre International.
- Reich, Hans H. (2000): Die Gegner des Herkunftssprachen-Unterrichts und ihre Argumente. In: Deutsch lernen, 25 (2): 112-126.
- Reich, Hans H. (2001): Tests und Sprachstandsmessungen bei Schülern und Schülerinnen, die Deutsch nicht als Muttersprache haben. In: Ursula Bredel u.a. (Hrsg.). Didaktik der deutschen Sprache. Ein Handbuch, 2. Teil. Paderborn: Schöningh, 914-923.
- Reich, Hans H. und Hans-Joachim Roth (2002): Spracherwerb zweisprachig aufwachsender Kinder und Jugendlicher. Ein Überblick über den Stand der nationalen und internationalen Forschung. Hamburg: Behörde für Bildung und Sport.
- Rösch, Heidi (2001): Zweisprachige Bildung in Berlin im Elementar- und Primarbereich. In: Essener Linguistische Skripte, 1 (1): 23-44.
- Rosenthal, Robert (1994): Parametric measures of effect size. In: Harris Cooper und Larry V. Hedges (Hrsg.). The handbook of research synthesis. New York: Russel Sage Foundation, 231-244.
- Rossell, Christine H. (1990): The effectiveness of educational alternatives for limited-English-proficient children. In: Gary Imhoff (Hrsg.). Learning in two languages. News Brunswick: Transaction Publishers, 71-121.
- Rossell, Christine H. (1992): Nothing matters? A critique of the Ramirez et al. longitudinal study of instructional programs for language minority children. In: Bilingual Research Journal, 16 (1&2): 159-186.
- Rossell, Christine H. (1998): Mystery on the bilingual express: A critique of the Thomas and Collier Study, <http://www.ceousa.org/collier.html>, READ Institute.
- Rossell, Christine H. (2000): The federal bilingual education program. In: Diane Ravitch (Hrsg.). Brookings papers on education policy. Washington: Brookings Institution Press, 215-243.
- Rossell, Christine H. (2003): The near end of bilingual education. In: Education Next (Fall): 44-52.
- Rossell, Christine H. und Keith Baker (1996): The educational effectiveness of bilingual education. In: Research in the Teaching of English, 30 (1): 7-74.
- Rossell, Christine H. und Julia Kuder (2005): Meta-murky: A rebuttal to recent meta-analyses of bilingual education. In: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI) (Hrsg.). The effectiveness of bilingual school programs for immigrant children. Berlin: Wissenschaftszentrum Berlin für Sozialforschung, Discussion paper SP IV 2005-601, 43-76.

- Roth, Hans-Joachim (2003): Bilinguale Alphabetisierung und die Entwicklung von Textkompetenz - am Beispiel einer portugiesisch-deutschen Klasse im zweiten Schuljahr. In: Zeitschrift für Erziehungswissenschaft, 6 (3): 378-402.
- Rothfarb, Sylvia H., Maria J. Ariza und Rafael Urrutia (1987): Evaluation of the Bilingual Curriculum Content (BCC) Project: A three-year study. Final report. Dade County, Office of Educational Accountability.
- Sachverständigenrat für Zuwanderung und Integration (2004): Migration und Integration - Erfahrungen nutzen, Neues wagen. Jahresgutachten 2004 des Sachverständigenrats für Zuwanderung und Integration. Berlin.
- Saldade, M., S. P. Mishra und M. Medina (1985): Bilingual instruction and academic achievement: A longitudinal study. In: Journal of Instructional Psychology, 12 (1): 24-30.
- Sandfuchs, Uwe und Clemens Zumhasch (2002): Wissenschaftliche Begleituntersuchung zum Schulversuch Deutsch-italienische Grundschule Wolfsburg - Reflexionen und ausgewählte Ergebnisse. In: Interkulturell (1/2): 104-139.
- Selinker, Larry (1992): Rediscovery interlanguage. London u.a.: Longman.
- Siebert-Ott, Gesa (1999): Zweisprachigkeit und Schulerfolg. Solingen: Landeszentrum für Zuwanderung NRW.
- Skoczylas, Rudolph V. (1972): An evaluation of some cognitive and affective aspects of a Spanish-English bilingual education program. Albuquerque: The University of New Mexico.
- Skolverket (2002): More languages - more opportunities: Skolverket, Offprint of report no. 228, <http://www.modersmal.net/eschola/summary.pdf>.
- Skutnabb-Kangas, Tove und Perri Toukomaa (1977): Teaching migrant children's mother-tongue and learning the language of the host country in the context of the socio-cultural situation of the migrant family (UNESCO-Report). Tampere: Forschungsbericht der Universität Tampere.
- Slavin, Robert E. und Alan Cheung (2005): A synthesis of research on language of reading instruction for English language learners. In: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI) (Hrsg.). The effectiveness of bilingual school programs for immigrant children. Berlin: Wissenschaftszentrum Berlin für Sozialforschung, Discussion paper SP IV 2005-601, 5-42.
- Slavin, Robert E. und Alan Cheung (2003): Effective reading programs for English language learners. A best-evidence synthesis: Center for Research on the Education of Students Placed At Risk (CRESPAR), Johns Hopkins University.
- Tannenbaum, Michal (2003): On mother tongue (And Fathers Too): Language and immigration from a developmental perspective. In: Newsletter International Society for the Study of Behavioural Development, Nr. 2: 11-13.
- Thomas, Wayne P. und Virginia Collier (1997): School effectiveness for language minority Students. Washington: National Clearinghouse for Bilingual Education.
- Thomas, Wayne P. und Virginia P. Collier (2002): A national study of school effectiveness for language minority students. Long-term academic achievement.
- Unabhängige Kommission „Zuwanderung“ (2001): Zuwanderung gestalten - Integration fördern. Berlin.
- U.S. Department of Education (2004): U.S. Education. General information, <http://www.ed.gov/about/offices/list/ous/international/usnei/us/edlite-structgeninfo.html> [15.12.2004].

- van Tubergen, Frank, Ineke Maas und Henk Flap (2004): The economic incorporation of immigrants in 18 Western societies: Origin, destination, and community effects. In: *American Sociological Review*, 69 (5): 704-727.
- Verhoeven, Ludo (1987): *Ethnic minority children acquiring literacy*. Dordrecht: Foris Publication.
- Verhoeven, Ludo (1994): Transfer in bilingual development: The linguistic interdependence hypothesis revisited. In: *Language Learning*, 44 (3): 381-415.
- Wagenaar, E. (1993): *Tweetaligheid in het aanvangsonderwijs. Een onderzoek naar de effecten van tweetalig kleuteronderwijs op de schoolloopbaan van Marokkaanse kinderen*. Amsterdam: Het Spinhuis.
- Willig, Ann (1985): A meta-analysis of selected studies on the effectiveness of bilingual education. In: *Review of Educational Research*, 55 (3): 269-317.
- Zickgraf, Arnd (2005): Anschluss im Sprachlernexpress. In: *Bildung PLUS*,  
[http://www.forumbildung.de/templates/imfokus\\_inhalt.php?artid=395](http://www.forumbildung.de/templates/imfokus_inhalt.php?artid=395) (Teil 1)  
[http://www.forumbildung.de/templates/imfokus\\_inhalt.php?artid=401](http://www.forumbildung.de/templates/imfokus_inhalt.php?artid=401) (Teil 2) [29.3.05].
- Zimmer, Dieter E. (1998): Lieber gleich ins kalte Wasser. Die zweisprachige Erziehung in den USA ist ein Lehrstück dafür, wie leicht sich vermeintlicher Fortschritt verrennt. In: *DIE ZEIT*, Nr. 48.

## **Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI)**

Die Arbeitsstelle wurde 2003 am Wissenschaftszentrum Berlin für Sozialforschung (WZB) eingerichtet und wird für drei Jahre aus Mitteln des Bundesministeriums für Bildung und Forschung finanziert. Sie ist mit zwei Wissenschaftlerinnenstellen ausgestattet und hat eine fünfköpfige Steuerungsgruppe sowie einen Beirat, der 14 Mitglieder aus Wissenschaft, Politik und Medien umfasst.

Maßgebend bei der Einrichtung des Projekts waren zwei Grundüberlegungen: Durch vielfältige und anhaltende Wanderungsbewegungen in Europa ist das Zusammenleben von Menschen unterschiedlicher kultureller Zugehörigkeit zur Normalität geworden. Diese Situation beinhaltet Herausforderungen und Konfliktpotenziale, aber auch Chancen. Zweitens geht das Projekt davon aus, dass in der sozialwissenschaftlichen Forschung sowohl weit entwickelte theoretische Modelle als auch eine umfangreiche empirische Evidenz für die Analyse von Problemen und die Unterstützung politischer Entscheidungen vorliegen, diese Ressourcen aber aus verschiedenen Gründen nicht optimal genutzt werden.

Ein zentrales Ziel der Arbeitsstelle ist die problemgeleitete systematische Bilanzierung der in verschiedenen wissenschaftlichen Disziplinen erarbeiteten Forschungsevidenz zu ausgewählten Fragestellungen im Themenfeld. Die Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI) will so zur Weiterentwicklung der Forschung beitragen sowie – soweit möglich – Grundlagen für politisches Handeln bieten. Allgemein macht es sich die AKI zur Aufgabe, die Kommunikation und Kooperation zwischen Wissenschaft, Politik und Öffentlichkeit im Themenfeld Migration-Integration-Konflikte zu fördern. Sie möchte auch einen Beitrag dazu leisten, dass das Forschungsfeld Interkulturelle Konflikte und gesellschaftliche Integration mehr Kontur und Gestalt gewinnt und wissenschaftlich wie gesellschaftlich stärker sichtbar wird.

Zu den Aktivitäten der AKI gehören neben der Erstellung der Forschungsbilanzen interdisziplinäre Expertenworkshops; außerdem gibt sie den AKI-Newsletter heraus.

Weitere Informationen zur Arbeitsstelle und ihren Veranstaltungen, der AKI-Newsletter sowie weitere Veröffentlichungen sind online unter [www.aki.wz-berlin.de](http://www.aki.wz-berlin.de) verfügbar. Der Newsletter kann als Print- oder Online-Ausgabe bei der Arbeitsstelle abonniert werden.

Die Arbeitsstelle am WZB: Dr. habil. Karen Schönwälder (Leiterin), Dipl. Soz. Janina Söhn (wissenschaftliche Mitarbeiterin), Manuela Ludwig (Sekretariat), Clémence Naaimi (studentische Mitarbeiterin).

Die Mitglieder der Steuerungsgruppe der AKI: Prof. Dr. Klaus J. Bade, Osnabrück, Prof. Dr. Hartmut Esser, Mannheim, Prof. Dr. Wilhelm Heitmeyer, Bielefeld, Prof. Dr. Amélie Mummendey, Jena, Prof. Dr. Friedhelm Neidhardt, Berlin.

Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI)  
Wissenschaftszentrum Berlin für Sozialforschung  
Reichpietschufer 50  
10785 Berlin  
Tel.: 030-25491-352  
[aki@wz-berlin.de](mailto:aki@wz-berlin.de)  
[www.aki.wz-berlin.de](http://www.aki.wz-berlin.de)